AD_____

Award Number:
W81XWH-07-1-0582

TITLE:
Differential Phosphoprotein Profiling of Tamoxifen Response

PRINCIPAL INVESTIGATOR:
Kolbrun Kristjansdottir, PhD

CONTRACTING ORGANIZATION:
University of Chicago
970 E 58th Street
Chicago, IL 60637

REPORT DATE:
August 2010

TYPE OF REPORT:
Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

x Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
08/29/2010	Final	30 Jul 2007 - 29 Jul 2010

4. TITLE AND SUBTITLE
Differential Phosphoprotein Profiling of Tamoxifen Response

5a. CONTRACT NUMBER
W81XWH-07-1-0582

5b. GRANT NUMBER
GRANT00205192

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kolbrun Kristjansdottir
Email: kolla@uchicago.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Chicago
970 E 58th Street
Chicago, IL 60637

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT - See Next Page.

15. SUBJECT TERMS
tamoxifen resistance, proteomics, biomarker, phosphorylation, estrogen receptor, breast cancer.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	47	**19b. TELEPHONE NUMBER** *(include area code)*

14. ABSTRACT

Advanced breast cancers that initially respond well to tamoxifen treatment eventually become refractory to this compound. Several mechanisms of acquired resistance have been hypothesized, including crosstalk between ER and growth factor receptor tyrosine kinase pathway. The cumulative data from clinical studies show that overexpression of HER-2 and/or EGFR, and high levels of phosphorylated Akt or ERK, contribute to tamoxifen resistance in some patients. HER-2, EGFR, Akt and ERK are all kinases and components of signaling pathways critical to cell growth and survival, highlighting the need for global phosphoproteome analysis. In this report I describe a method for comparison of global phosphoprotein profiles involving stable isotope labeling, a phosphoprotein affinity step, 1-D SDS-PAGE and LC-MS/MS. I applied this method, differential phosphoprotein profiling to compare phosphoprotein profiles in MCF-7 (tamoxifen sensitive) and MCF-7/HER2-18 (tamoxifen resistant) cells and to examine their regulation by tamoxifen. I found that FADD and other proteins involved in apoptosis were identified in the phosphoenriched fraction of MCF-7 cells but not MCF-7/HER2-18 cells. I also found several proteins regulated by tamoxifen. For example, phosphorylation of XRCC1 on XXX is decreased in MCF-7/HER2-18 cells but not in MCF-7 cells. Both FADD and XRCC1 have previously been described as being involved in tamoxifen resistance showing that phosphoprotein profiling is a feasible method for identifying proteins relevant to tamoxifen resistance.

Table of Contents

Introduction

Breast cancer remains the most common malignancy affecting women in the United States. About 80% of breast cancers are estrogen-receptor-alpha-positive (ERα+), some of which respond to estrogen hormone therapy. ERα is a ligand-activated transcription factor that plays a critical role in the etiology of breast cancer [1-3]. Selective estrogen receptor modulators (SERMs) have variable agonistic and/or antagonistic activities, depending on the type of ER (α versus β), tissue context, and interactions with different proteins such as transcriptional co-activator or co-repressors [4]. The first SERM, tamoxifen, revolutionized breast cancer treatment when it came into use some three decades ago. In ERα breast cancer cells, tamoxifen blocks cancer growth by competing for binding to ER and cuts recurrence risk in half [5] [6]. More recently, tamoxifen has been shown to prevent breast cancer in high-risk women [7] [8]. Even in patients with ERα-positive breast cancer, only 40–50% of patients benefit from tamoxifen treatment, suggesting that a substantial fraction of ER-positive cancers are resistant to this drug. Additionally, advanced breast cancers that initially respond well to tamoxifen eventually become refractory to this compound. In some cases, tamoxifen can even act as a growth stimulatory signal. Several mechanisms of resistance have been hypothesized, including crosstalk between ER and other proliferative signals, such as growth factor receptor tyrosine kinase pathways [9-12]. The cumulative data from clinical studies show that overexpression of HER-2 and/or EGFR, and high levels of phosphorylated Akt or ERK, contribute to tamoxifen resistance in some patients [13-16]. HER-2, EGFR, Akt and ERK are all kinases and components of signaling pathways critical to cell growth and survival, highlighting the need for global phosphoproteome analysis.

Although many biomarkers for breast cancer prognosis and therapy initially appeared attractive, over the years most of them have failed to become clinically useful, with the exception of hormone receptors (ER and PR) and the HER-2 tyrosine kinase receptor [17, 18]. Although ER status provides prognostic information, the major clinical value is to assess the likelihood that a patient will respond to endocrine therapy [2, 19]. HER2 is overexpressed in 25 to 30 percent of breast cancers, increasing the aggressiveness of the tumor [20]. The drug Trastuzumab (Herceptin) is a monoclonal antibody directed against the HER-2 and has a survival benefit when combined with chemotherapy in patients with metastatic breast cancer that overexpress HER-2 [21]. However, tumors that overexpress HER2 tend to be ERα negative and thus represent a separate treatment group. Current prognostic classifications are thus not enough to represent the broad clinical heterogeneity of breast cancer, making it difficult to target therapeutic strategies to each patient. A major component of prognosis for patients undergoing endocrine therapy is the acquired resistance to tamoxifen. Finding biomarkers for tamoxifen resistance and/or drugs that could help overcome the resistance is a very important topic.

New reporters that could be used in combination with existing markers for screening of breast cancer cells for treatment decisions or to predict therapy outcome are still needed. A major component of prognosis for patients undergoing endocrine therapy is the acquired resistance to tamoxifen. Finding reporters for tamoxifen

resistance and/or drugs that could help overcome the resistance is a very important topic.

Thanks to recent advances in technology and the ability to analyze enormous amounts of data, proteomics is poised to have a significant effect on cancer research. Although gene expression patterns of cancerous cells have been extensively studied, there is a dearth of information on protein expression and protein modification patterns. This is important because gene expression alone cannot determine the activation state of cellular proliferation signaling pathways. Aberrations in the regulation of these pathways are a key to the development and progression of cancers. The activity of signaling proteins depends on their interactions with other proteins and modifications (phosphorylations) they undergo over time, areas that proteomics is able to address [22, 23].

Before starting this project, I had developed and published a method for enrichment of phosphoproteins [24]. The methodology involves a phosphoprotein affinity step, 1-dimensional SDS-PAGE and ESI LC-MS/MS and is termed PA-GeLC-MS/MS. By combining the phosphoprotein enrichment method with stable isotope labeling relative quantitation of phosphoprotein profiles can be obtained. I refer to this combined method as differential phosphoprotein profiling. The overall goal of this project is obtain global phosphoprotein profiles of tamoxifen response and to compare responses in tamoxifen sensitive and resistant cell lines to identify markers of tamoxifen response. In this final report I describe phosphoprotein profiling of MCF-7 (tamoxifen sensitive) and MCF-7/HER2-18 (tamoxifen resistant) cells and report several proteins that respond differently to tamoxifen treatment in these two cell lines.

Body

PHOSPHOPROTEIN ENRICHMENT FROM CONTROL AND TAMOXIFEN TREATED MCF-7 AND MCF-7/HER2-18 CELLS

Differential phosphoprotein profiling was performed on two cell lines. First, the MCF-7 breast cancer cell line is estrogen receptor positive, responds to estrogen stimulation and is sensitive to tamoxifen. Several cell lines have been generated that are resistant to tamoxifen treatment. As mentioned previously, overexpression of HER2 has been described in patients with acquired tamoxifen resistance [29]. The tamoxifen resistant cell line used in these experiments, MCF-7/HER2-18, was generated by overexpressing full-length HER2 kinase in MCF-7 cells. The authors tested for response to tamoxifen by implanting MCF-7/HER2-18 or MCF-7 control cells into nude mice. Both cells only produced tumors when stimulated with estrogen, but MCF-7/HER2-18 grew much more rapidly. Tamoxifen inhibited growth in the MCF-7-derived tumors but not in the MCF-7/HER2-18 derived tumors [20].

Phosphoprotein enrichment experiments were performed on both MCF-7 (tamoxifen sensitive) cells and MCF-7/HER2-18 (tamoxifen resistant) cells (Figure 1, see next page). The cells were SILAC labeled with DMEM-Flex media (Invitrogen) without phenol red and contained high glucose (4500 mg/ml), 1mM sodium pyruvate, 10% heat-inactivated dialyzed fetal bovine serum, 1% penicillin/streptomycin and 0.3 mg/ml L-glutamine. Briefly, two equal amounts of cells were seeded onto plates, one was grown in “light” (L-lysine and L-Arginine) and the other in “heavy” ($^{13}C_6$ L-lysine and $^{13}C_6{}^{15}N_4$ L-Arginine) media for >10 doublings.

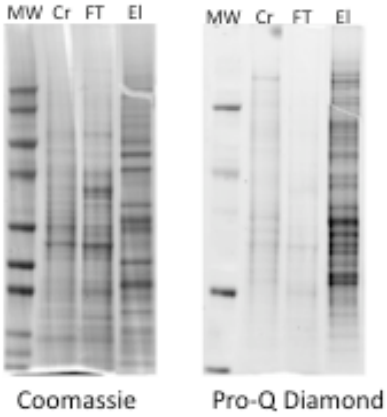

Figure 2. Phosphoprotein enrichment of proteins from MCF-7/HER2-18 cells. MCF-7/HER2-18 cells were split into two equal samples and grown in either heavy or light SILAC media. The heavy cells were then treated with 10 nM Tamoxifen and the light cells with ethanol as control, for a total of 30 minutes. The samples were lysed and mixed at 1:1. Phosphoproteins were isolated using a phosphoaffinity column (Pro-Q Diamond, Invitrogen/Molecular Probes). Lysate (L), flowthough (FL) and Eluate (E) from the phosphoaffinity column were subjected to SDS-PAGE and the gel stained with Imperial Coomassie to visualize proteins and Pro-Q Diamond fluorescent stain to visualize phosphoproteins. Representative figure for MCF-7/HER2-18 and MCF-7 cells.

Prior to treatment cells were serum starved for 2 hours. The cells were then treated for 30 minutes with 10 nM 4-hydroxy-tamoxifen (Sigma) or ethanol as control. Whole cell lysates were prepared from 7 x 10^7 cells in 1.5 ml of lysis buffer (ProQ lysis buffer with 1 µM sodium fluoride, 1 µM okadaic acid and 0.1 µM sodium orthovanadate). The supernatant was collected, and protein yields were determined by Bradford analysis using Bio-Rad protein assay reagent. About 5 mg of lysate was obtained from each sample. A sample of the lysate was stored for follow-up analysis using Western blots. 2.5 mg of lysate from light cells and 2.5 mg of lysate from heavy cells was mixed and the combined lysate was loaded onto pre-equilibrated Pro-Q Diamond resin, the column washed and phosphoproteins eluted. The lysate, flow-through and eluate were

concentrated in 10 kDa MWCO Vivaspin concentrators at 4 °C and washed with 50 mM Tris, pH 7.5. The samples were mixed with Laemmli buffer and incubated at 95°C for 5 min before loading on NuPAGE 2-12% gradient gels. The gel was stained for phosphoproteins using Pro-Q Diamond stain and subsequently for proteins with Imperial Coomassie stain. Coomassie stained protein was visible in all three fractions including the flow through (Figure 2, see previous page). The dark staining in the eluate fraction

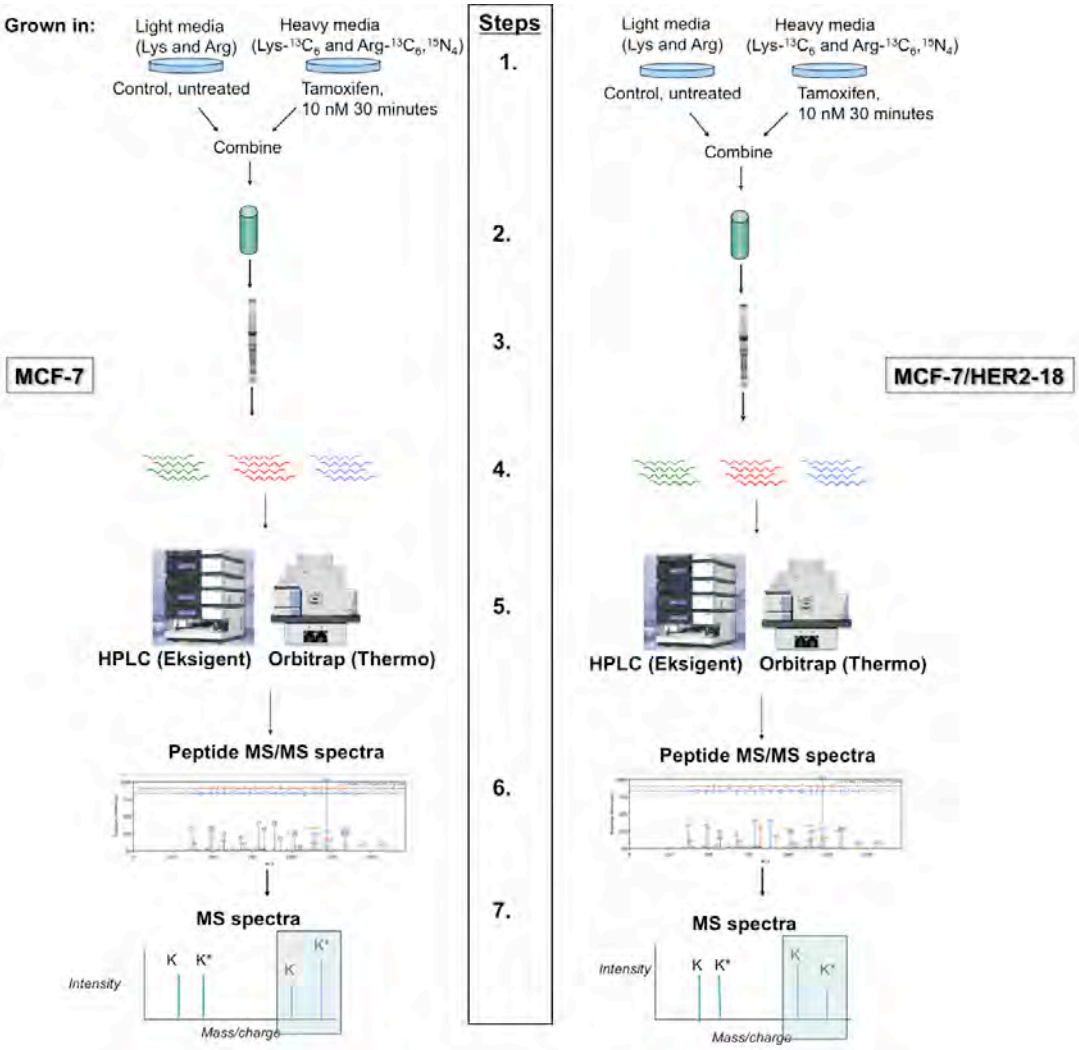

Figure 1. Scheme for differential phosphoprotein profiling. Two cell lines were used for analysis, MCF-7 and MCF-7/HER2-18. (1) One sample is grown in media with stable isotope labeled arginine (Arg) and lysine (Lys) (heavy sample) and another grown in regular media (light sample). Heavy sample is treated with 10 nM Tamoxifen for 30 minutes, the light sample is untreated control. Samples are then combined, subjected to (2) phosphoenrichment (Pro-Q Diamond resin, Invitrogen/Molecular Probes), separation by (3) SDS-PAGE (cut into 18 sections). The samples are then (4) digested and peptides extracted and subjected to (5) reversed phase nanoLC-MS/MS. Peptide and protein identification from (6) MS/MS spectra using Mascot, X!Tandem and compiled in Scaffold. Relative abundance calculated from MS spectra (7) using XPRESS in CPAS. Experiment was repeated identically except tamoxifen treatment was performed on the light sample (gel B in Table 1). Peptides whose abundance ratios differ between MCF-7 and MCF-7/HER2-18, represented by blue peptide in shadowed box, are the ones of interest.

and the scarcity of phosphoproteins in the flowthrough fraction shows that the Pro-Q Diamond resin selectively binds phosphoproteins.

MASS SPECTROMETRY OF THE ENRICHED PHOSPHOPROTEINS

Proteins were extracted for mass spectrometry analysis from the ProQ elution gel lane of the SDS-PAGE gel (Figure 2, elution lane). Briefly, the molecular weight region above 10 kD was divided into 20 sections, about 0.5 cm each. The top two and second two sections were combined, giving a total of 18 sections. Each section was cut into small pieces, each ~1 mm^3. Sections were washed in water and completely destained using 100 mM ammonium bicarbonate in 50% acetonitrile. A reduction step was performed by addition of 100 µl of 50 mM ammonium bicarbonate pH 8.9 and 10ul of 10 µM TCEP and allowed to reduce in 37 ºC for 30 min. The proteins were alkylated by adding 100 µl of 50 mM iodoacetamide and allowed to react in the dark for 40 min. Gel sections were washed in water, initially dried with acetonitrile followed by a SpeedVac step of 30 min. Digestion was carried out using sequencing grade modified trypsin (40 ng/ml, Promega) in 50 mM ammonium bicarbonate. Sufficient trypsin solution was added to swell the gel pieces, which were kept in 4º C for 45 min and then incubated at 37º C overnight. Sections containing proteins larger than 150 kD were pre-digested with Lys-C (0.25 mg/ml, Princeton Separations) in 6-8 M Urea overnight at 25 ºC, diluted to final concentration of less than 2 M Urea then digested with trypsin as described above. Peptides were extracted from the gel pieces with 5% formic acid.

All mass spectrometry was performed in the Mayo Proteomics Research Center, on Thermo LTQ-Orbitrap Hybrid FT Mass Spectrometers. The peptide samples were loaded to a 0.25 µl C8 trapping cartridge OptiPak custom-packed with Michrom BioResources Magic C8, 5 µm, 200A, washed, then switched in-line with a 20 cm by 75 um C18 'packed spray tip' nano column packed with Magic C18AQ, 5 µm, 200A, for a 2-step gradient, where mobile phase A is water/acetonitrile/formic acid 98/2/0.2 and mobile phase B is acetonitrile/isopropanol/water/formic acid 80/10/10/0.2. Using a flow rate of 350 nl/min, a 90 min, 2-step LC gradient was run from 5% B to 50% B in 60 min, followed by 50%-95% B over the next 10 min, hold 10 min at 95% B, back to starting conditions and re-equilibrated. The samples were analyzed via electrospray tandem mass spectrometry (LC-MS/MS) on the LTQ-Orbitrap using a 60,000 RP Orbi survey scan, m/z 375-1950, with lock masses, followed by 5 LTQ CAD scans with isolation width of 1.6 Da on doubly and triply charged-only precursors between 375 Da and 1500 Da. Ions selected for MS/MS were placed on an exclusion list for 60 s using low mass exclusion of 1.0 Da, high mass exclusion of 1.6 Da.

The mass spectrometry data were converted to .mgf files via .mzXML intermediates and searched using Mascot using the SILAC (MD) quantitation parameter. A fragment ion mass tolerance of 50 ppm and a parent ion tolerance of 0.6 Da were specified. Oxidation of methionine, phosphorylation (S, T, Y) and carbamidomethyl (C) were specified as variable modifications. Mascot results were loaded into Scaffold (Proteome Software), which uses Peptide and Protein prophet to calculate probabilities. Scaffold also conducted an X!Tandem search using the parameters used for Mascot.

Comparative Proteomics Analysis System(CPAS) is a open-source analytic system based on the modules developed in the Trans Proteomic Pipeline from Institute

of Systems Biology (Seattle) [30]. CPAS was used to perform quantitation on the data from mzXML files. The analysis pipeline involved performing X!Tandem searches (using the parameters described above), converting the results to .pepXML format, processing by Peptide Prophet for statistical evaluation of peptide identifications and Xpress software for relative peptide quantification. The peptide results from all 18 sections were exported and combined into one excel file. Proteins were compiled and protein averages calculated using a Perl script provided by the Hanash lab at Fred Hutch (Seattle). Experiments were performed in duplicate, gel A where heavy cells treated with tamoxifen and light were untreated and gel B where light cells were treated with tamoxifen and heavy were untreated (Table 1).

Table 1. Overview of mass spectrometry experiments.

SILAC labeled

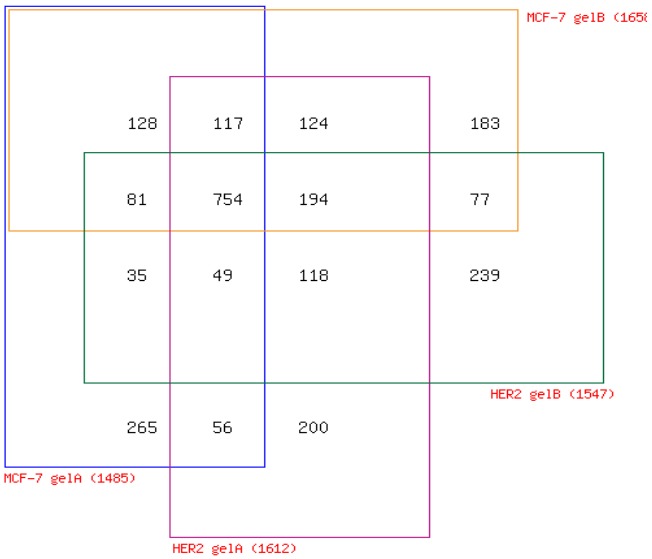

Figure 3. Venn Diagram shows overlap between proteins identified from differential phosphoprotein profiling. The Venn diagram shows two replicates from the MCF-7 cell line (MCF-7 gel A and B) and two replicates from the MCF-7/HER2-18 cell line (HER2 gel A and B). The diagram was made in

Cells	Name	Tamoxifen treatment	Control	# Sections	Status
MCF-7	GelA	Light	Heavy	18	Completed
MCF-7	GelB	Heavy	Light	18	Completed
MCF-7 /HER2-18	GelA	Light	Heavy	18	Completed
MCF-7 /HER2-18	GelB	Light	Heavy	18	Completed

RESULTS

PHOSPHOPROTEIN PROFILING OF MCF-7 CELLS WITH AND WITHOUT TAMOXIFEN TREATMENT

Using these methods over 1400 proteins were identified from the Pro-Q Diamond enriched fraction of MCF-7 cells (Figures 1 and 2). Specifically, a protein probability of >99%, peptide probability of >95% and a minimum of 2 unique peptides per protein identification were required in Scaffold giving a 5.4% false discovery rate (FDR) for

peptides and 0.1% FDR for proteins. Experiments were performed in duplicate: In experiment A heavy cells were treated with tamoxifen and light cells were untreated and experiment B light cells were treated with tamoxifen and heavy were untreated (Table 1). The two replicates had 1080 identical proteins from a total of 1483 or 73% protein overlap (Figure 3). When similar proteins are included, as measured by GeneGo software, the number goes up to 83%. Only identical proteins identified in both samples were used for further analysis. Quantitative analysis will reveal which proteins are affected by tamoxifen treatment (see below).

THE EFFECT OF TAMOXIFEN ON THE PHOSPHOPROTEOME OF MCF-7 CELLS

Quantitation was performed as described above and only protein ratios with less than 10% standard deviation between gelA and gelB (Table 1) were averaged and included in further analysis. The vast majority of proteins did not change substantially in abundance. About 20 proteins were identified that decreased >25% and about 30 proteins that increased >25% in the tamoxifen treated sample. Gene ontology analysis of these proteins reveals that they are involved in several important processes such as protein transport, DNA repair, signal transduction and protein biosynthesis.

PHOSPHOPROTEIN PROFILING OF MCF-7/HER2-18 CELLS WITH AND WITHOUT TAMOXIFEN TREATMENT

Phosphoprotein profiling on MCF-7/HER2-18 tamoxifen resistant cells resulted in identification of over 1500 proteins (protein probability >99%, peptide probability >95%, requiring a minimum of 2 unique peptides per protein identification). Among the proteins identified were HER2 kinase, as expected since it is over-expressed in the cell line. HER-2 protein coverage was 36%. Experiments were performed in duplicate: In experiment A heavy cells were treated with tamoxifen and light cells were untreated and experiment B light cells were treated with tamoxifen and heavy were untreated (Table 1). The two replicates had 1115 identical proteins from a total of 1547 or 72% protein overlap (Figure 3). Only identical proteins identified in both samples were used for further analysis. Quantitative analysis will reveal which proteins are affected by tamoxifen treatment (see below).

THE EFFECT OF TAMOXIFEN ON THE PHOSPHOPROTEOME OF MCF-7/HER2-18 CELLS

Quantitation revealed that the vast majority of proteins did not change substantially in abundance. 5 proteins were identified that decreased >25% in the tamoxifen treated sample and 8 proteins that increased >25% in the tamoxifen treated sample. Gene ontology analysis of these proteins reveals that the proteins are involved in several important processes such as DNA repair, protein transport and signal transduction.

COMPARING THE IDENTIFIED PROTEINS AND PHOSPHORYLATION SITES TO DATABASES OF PHOSPHORYLATION

Examination of proteins identified revealed that several phosphorylation sites were identified (Figure 4). The majority of phosphorylation sites (with the exception of MCF-7/HER2-18) are from similar or common proteins. Quantitative analysis will reveal which phosphorylation sites are affected by tamoxifen treatment (see below). All peptides contained phosphoserines and/or phosphothreonines. No phosphotyrosine containing peptides were detected. PhosphoELM is a database of phosphopeptides identified by mass spectrometry (Diella et al, 2008). The database contains 4078 protein sequences containing 16470 total phosphorylation sites (12025 (73%) phosphoserine, 2362 (14%) phosphothreonine and 2083 (13%) phosphotyrosine).

Overall, 38% of identified proteins, with identified phosphosites and without, identified from the MCF-7 cell line and 40% of proteins identified from MCF-7/HER2-18 were found in phosphoELM. As expected, the phosphorylation sites included all three types of phosphorylated amino acids (serine, threonine and tyrosine). The frequency of phosphoserine in the Pro-Q Diamond enriched proteins, >70%, corresponds nicely with the frequency of phosphoserine in the PhosphoELM database. 10% of the Pro-Q Diamond enriched proteins contain only phosphotyrosine sites in the PhosphoELM database. Again, the frequency correlates well with PhosphoELM database as a whole and indicates that the Pro-Q Diamond resin is not biased towards any of the phosphorylated residues.

COMPARISON OF IDENTIFIED PROTEINS IN MCF-7 AND MCF-7/HER2-18 CELLS

Proteins identified from all 4 experiments: MCF-7 cells (two experimental replicates A and B) and MCF-7/HER2-18 cells (two experimental replicates A and B) are compared in a Venn diagram in Figure 3. A significant overlap exists between the two cell lines, as can be expected, since the MCF-7/HER2-18 cell line was generated by overexpressing HER-2 in an MCF-7 cell line. Interestingly, several proteins were identified in only one cell line. Specifically, 128 proteins were found in both MCF-7 experiments but in neither HER-2 experiments and 118 proteins were found in both HER-2 experiments but in neither MCF-7 experiment. Analysis of these proteins using GeneGo revealed an enrichment of apoptotic molecules in the MCF-7/HER2-18 cell line This is of great interest to me and I will follow up on this interesting observation. I have

Figure 4. A comparison of phosphoeptides identified using differential phosphoprotein profiling. Phosphopeptides from MCF-7 Gel A (orange), MCF-7 Gel B (blue) , MCF-7/HER2-18 Gel A (red) and MCF-7/HER2-18 Gel B (green) were compared at the protein level using GeneGo software.

chosen FADD, an apoptotic adaptor molecule that recruits activated Caspase 8 or 10 to activated Fas and TNFR-1 (Tumor Necrosis Factor) receptors to follow up on. Studies have shown that inhibiting or reducing phosphorylation of Ser194 in the FADD protein in MCF-7 cells results in decreased sensitivity to tamoxifen treatment [37].

FADD was detected only in the phosphoenriched fraction from the MCF-7/HER2-18 cell line. However, the lack of detection in any mass spectrometry experiment does not necessarily mean that the protein is not there. Thus we analyzed FADD levels using RT-PCR and Western blots. The RT-PCR showed no chance in mRNA. Western blots showed similar amounts of FADD present in both MCF-7 and MCF-7/HER2-18 cell extracts (Figure 6). Thus, although FADD was not detected in the phosphoenriched fraction of MCF-7/HER2-18 it is not due to the protein being absent. Thus it is likely that FADD is phosphorylated in MCF-7/HER2-18 cells and not in MCF-7 cells. I was not able to identify the phosphorylation site by mass spectrometry or detect signal using anti-FADD phosphoSer194 antibody. Previously, it has been shown that phosphorylation of FADD on Serine194 is statistically different between breast tumor epithelial cells and matched undissected breast tissue [36]. I propose that phosphorylation of FADD on Serine 194 could be a marker for tamoxifen treatment efficacy.

THE EFFECT OF TAMOXIFEN ON THE PHOSPHOPROTEOME OF MCF-7 AND MCF-7/HER2-18 CELLS

I have compared the results from MCF-7 to MCF-7/HER2-18 phosphoprotein profiling of tamoxifen response and identified 26 proteins that respond to tamoxifen differently. All but three of these proteins are known to be phosphorylated and at least one of the three proteins is known to bind to a phosphoprotein and could thus have been purified on the Pro-Q Diamond resin as a phosphoprotein complex. Of these proteins, XRCC1 is a promising marker. A relationship has been shown between *XRCC1* polymorphisms and breast cancer risk that reported an inverse association between the Trp194 carriers and breast cancer risk (Patel et al, 2005). In particular, *XRCC1* Arg194Trp and Arg399Gln polymorphisms have been shown to affect *XRCC1* protein-product expression and to alter BER capacity.

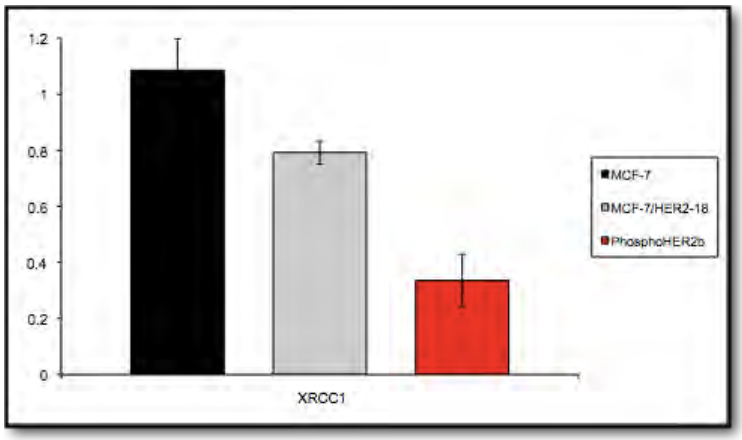

Figure 5. Tamoxifen treatment results in decreased phosphorylation of XRCC1 on Ser447/Thr453 in MCF-7/HER2-18 cells. Black bars show the duplicate MCF-7 experiments with ratios around 1:1 with and without tamoxifen treatment. Grey bars show the duplicate MCF-7/HER2-18 experiments showing a 20% decrease in XRCC1 levels after tamoxifen treatment. In particular, the phosphopeptide from XRCC1 containing pSer447/pThr453 decreased 70% after tamoxifen treatment (red bars).

I show in this report that two known phosphorylation

sites in XRCC1, Ser447 and Thr453, are detected in the tamoxifen resistant cell line and the levels of these significantly decreased after tamoxifen treatment (Figure 5). No antibody is available for this phosphorylation site but we did perform RT-PCR and saw a slight decrease in XRCC1 levels in response to tamoxifen in MCF-7/HER2-18 cells but not enough to explain the decrease in phosphorylation (Figure 6). How phosphorylation on Ser447 and Thr453 affects the function of XRCC1 is not clear. The kinase that phosphorylates Ser447 and Thr453 in XRCC1 is not known. Taken together, several potential markers for tamoxifen response have been identified from a single proteomic screen, showing the strength of this approach.

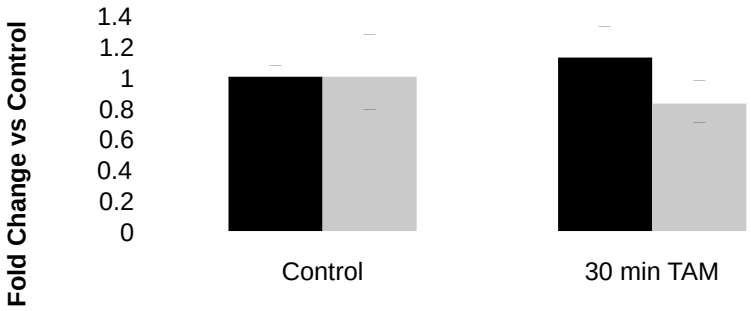

Figure 6. RT-PCR results showing no significant change in mRNA levels after tamoxifen treatment. MCF-7 cells (black bars), MCF-7/HER2-18 cells (grey bars).

Methods for RT-PCR and Western blots:

RT-PCR

Cells were seeded onto 6-well plates, treated with serum stripped media for 24 hours and then with the 10nM Estradiol or 10 nM Tamoxifen or equal volume ethanol as control for the indicated times. RNA was extracted using 1 ml ice-cold Trizol (Invitrogen) for 10 minutes and frozen at -80°C until ready for analysis. 1.2 µg total RNA was treated with Amplification Grade DNAse I (Invitrogen). cDNA was synthesized from one-half of the RNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). qPCR was performed in a 384-well plate on an ABI 7900HT (Applied Biosystems) using Power SYBR Green PCR Master Mix (Applied Biosystems) in a 5 µl reaction volume containing 2 µl of 1:40 diluted cDNA and 0.5 µl of 100 µM primers. PCR primers, designed using Primer Blast (www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi), were as follows: TNF-f: GCC AGA GGG CTG ATT AGA GA TNF-r: TCA GCC TCT TCT CCT TCC TG, IKBA-f: GATCCGCCAGGTGAAGGG, IKBA-R: GCAATTTCTGGCTGGTTGG, FOS, CHK2, RIPK1, PARP1-f: CAA CTT TGC TGG GAT CCT GT, PARP1-r: GGT CCC AAG AGG AAC GTC TA, EGR1-f: GCAAGTACCCCAACCGGC, EGR1-r: GCAAACTTCCTCCCACAAATGT, GAPDH-f: TGCACCACCAACTGCTTAGC, GAPDH-r: GGCATGGACTGTGGTCATGAG. QuantiTect primers for GAPDH, FADD, BET1, XRCC1, DTYMK, API5, PBK, PAK1, NUP62 and GGA1 were obtained from Qiagen: Hs_GAPDH_2_SG, Hs_FADD_1_SG, Hs_BET1_1_SG, Hs_XRCC1_1_SG, Hs_DTYMK_2_SG, Hs_API5_1_SG, Hs_PBK_1_SG, Hs_PAK1_1_SG, Hs_NUP62_2_SG, Hs_GGA1_1_SG. Fold-change calculations were performed using the comparative Ct method, using GAPDH as the endogenous control.

Western Blots

The cells were seeded onto 10 cm plates, treated with stripped media for >24 hours, then treated with 10 nM E2, TAM or ethanol as a control and incubated for the indicated times. Cells were then rinsed with PBS, scraped off of plate, spun down and the pellet frozen at -80°C until use. The pellet was lysed in ProQ lysis buffer with HALT phosphatase inhibitors and HALT protease inhibitors and mixed with 4X Laeemli buffer, boiled for 5 min and spun down. 50 µg of protein lysate was separated by SDS-PAGE on 4-12% NuPAGE gradient gels. The proteins were transferred to nitrocellulose membrane and blocked in 5% milk in TBST for 1 hour. The membrane was incubated with antibodies at 1:1000 dilutions at 4°C overnight. Primary antibodies were the following: αFADD, αFADD(pSer194), from Cell Signaling. The membrane was washed in TBST, and incubated with 1:5000 diluted secondary antibody (αRabbit from GE Healthcare) for 1 hour before addition of ECL reagent and developing film.

^{18}O labeled

MCF-7 /HER2-18	GelA	Light	Heavy	18	Completed
MCF-7 /HER2-18	GelB	Heavy	Light	18	In progress

To identify the effects of longer tamoxifen treatment I performed another proteomic experiment in MCF-7/HER-2 cells (listed as ^{18}O labeled in Table 1). Instead of using SILAC labeling, I tested an alternative, termed ^{18}O labeling. The benefits of ^{18}O labeling include removing the requirement for growth in labeled media. This will allow me to label patient samples in the future. Along with Don Wolfgeher in the lab, I optimized the 18O labeling protocol. We also had to generate in-house software for ^{18}O quantiation which was done by Jonathon Goya, in the lab.

Briefly, MCF-7/HER2-18 cells were maintained in DMEM media without phenol red and containing high glucose (4500 mg/ml), 1mM sodium pyruvate, 10% fetal bovine serum, 1% penicillin/streptomycin and 0.3 mg/ml L-glutamine and 0.1 mg/ml G418. Two equal amounts of cells were seeded onto plates and incubated in the same media as above except media was used that contained charcoal stripped serum. After 24 hours, the cells were then treated for 24 hours with 10 nM 4-hydroxy-tamoxifen (Sigma) or ethanol as control. Whole cell lysates, Pro-Q phosphoprotein enrichment, SDS-PAGE, trypsin digestion from gel were all performed as described above. The peptide samples were then spun down to dryness and then reconstituted in either 30 µL regular water or $H_2{}^{18}O$ (99%, Cambridge Isotope labs) and dry magnetic trypsin beads added to the solution for 24 hours at 37C. The sample was then spun down to dryness again. Right before mass spectrometry analysis the sample was resuspended in 30 µL water/AcN/formic acid and mixed at 1:1 ratio. Data analysis was performed as described above except the quantitation was performed with in-house software. The replicate experiment is awaiting mass spectrometry analysis, I expect the data next week. Since my criteria for accepting protein identifications is identification in two replicate samples, I cannot report the results for this analysis in this report.

COMPUTATIONAL SOLUTIONS TO COMPLEX SIGNALING ANALYSIS

One major problem I have encountered during this work is the low quality of proteomic software. Analysis takes very long and often includes manual validation of spectra and quantitation levels. To circumvent this problem I formed a collaboration with a computer scientist in the Computation Institute here at the University of Chicago, Sam Volchenboum. Our software takes advantage of the fact that once samples have been labeled with stable isotope and is mixed with an unlabeled sample, each peptide appears as a doublet (light, unlabeled and heavy, stable isotope labeled). This distinguishes peptides from background peaks and aids in the identification of peptides. In addition, since the isotope is added to the C-terminal of the peptide (in SILAC and $^{18}0$ labeling), C-terminal fragment ions (y-ions) are shifted between the two fragmentation spectra (light and heavy forms) from non-labeled and non-shifted N-terminal fragment ions (b-ions). Utilizing this information we developed a fast and reliable method for automated validation of Mascot search results from high accuracy mass spectrometry data. We can identify isotopic pairs within searched Mascot data (DAT file), and these pairs represent the highest confidence peptide matches. Our software, termed Validator, demonstrated a false discovery rate of only 2% while retaining most high-Mascot scoring peptides and eliminating most low-scoring ones. We also demonstrated that our software identifies peptide pairs based only on their difference in precursor mass owing to the presence of the stable isotope label using no Mascot-specific information. We were able to corroborate 81% of identified peptide pairs using conventional database search engines and published the paper in Journal of Molecular and Cellular Proteomics [38] (the paper in its entirety is found in the appendix). We are currently working on a second publication; describing a program we have termed Identifier. Identifier takes the proteome of an organism, for example yeast, and generates in silico digested peptides listing the peptide sequence, mass and the identity and mass of b- and y-ions. Thus the workflow will involve using Validator to identify peptide pairs from the raw data and comparing the mass and fragmentation patterns of peptides to the in silico digested proteome. This allows for very rapid analysis of mass spectrometry data and represents a novel method of protein identification that can be used instead of or in addition to conventional database search engine methods. Finally, Jonathan Goya in our lab has written a quantitation module that will be added to our software and allow for complete analysis of proteomic samples in a rapid, reliable manner. The quantitation module will be published as a separate paper and the manuscript is in preparation.

Key Research Accomplishments

- I have performed phosphoprotein enrichment from tamoxifen treated and control untreated samples from tamoxifen sensitive (MCF-7) and tamoxifen resistant (MCF-7/HER2-18) cell lines. The experiment was performed twice for each cell line.
- Each experiment identified over 1400 proteins and dozens of phosphorylation sites were identified.
- I have compared the results from MCF-7 to MCF-7/HER2-18 phosphoprotein profiling of tamoxifen response. In particular, XRCC1 is a promising marker for tamoxifen resistance. I show in this report that two known phosphorylation sites in XRCC1, Ser447 and Thr453, are detected in the tamoxifen resistant cell line and the levels of these significantly decreased after tamoxifen treatment
- In addition, examining proteins only found in the phosphoenriched section of one of the cell lines revealed an abundance of proteins involved in apoptosis. One of these proteins, FADD, has previously been shown to result in resistance to tamoxifen when phosphorylation on Ser194 is blocked. We found that FADD was present in phosphoenriched fraction from MCF-7 cells but was not detected in MCF-7/HER2-18 cells. This is not due to changes in proteins amounts since the RT-PCR showed no chance in mRNA and Western blots showed similar amounts of FADD present in both MCF-7 and MCF-7/HER2-18 cell extracts.
- In collaboration with Sam Volchenboum, Instructor in Pediatrics and the Computational Institute at the University of Chicago, I developed a fast and reliable method for automated validation of Mascot search results from high accuracy mass spectrometry data which was published the paper in Journal of Molecular and Cellular Proteomics.
- A quantitation module for stable isotope labeled proteomic data analysis was written in collaboration with Jonathan Goya in lab.

Reportable Outcomes

A. Talks and poster presentations

1. **Cancer Biology Training Consortium, Chairs and Program Directors Retreat and Annual Meeting (CABTRAC)** in Basin Harbor Resort, Vermont, September 30th-October 2 2007. Presented poster entitled: "Phosphoprotein profiling for quantitative analysis of phosphorylated proteins"
2. **American Association for Cancer Research (AACR) Annual Meeting.** San Diego, California, April 10-15th, 2008. Presented poster entitled: "Differential phosphoprotein profiling of Tamoxifen response".
3. **Department of Defense (DOD) Breast Cancer Research Program (BCRP) Era of Hope 2008 Meeting** in Baltimore, MD in June 25-28th, 2008. Presented poster entitled: "Differential phosphoprotein profiling of Tamoxifen response".
4. **University of Chicago Annual Molecular Biosciences Retreat**, Galena, IL November 7-9, 2008. Oral presentation titled: "Differential Phosphoprotein Proteome Profiling of Tamoxifen Response"
5. **29th Annual Minisymposium on Reproductive Biology.** Evanston, IL, October 6th, 2008. **Presented poster entitled:** Presented poster entitled: "Differential phosphoprotein profiling of Tamoxifen response".
6. **Midwest Breast Cancer Research Symposium.** Iowa City, Iowa. July 17-19th, 2009. Presented poster entitled: "Differential phosphoprotein profiling of Tamoxifen response".
7. **Gordon Conference: Hormone Action In Development & Cancer.** Holderness, NH, July 26-31st, 2009. Presented poster entitled: "Differential phosphoprotein profiling of Tamoxifen response".
8. **University of Chicago Department of Molecular Genetics and Cell Biology Miniretreat.** Chicago, IL March 11th, 2010. "Differential Phosphoprotein Proteome Profiling of Tamoxifen Response"**.**

B. Publications and manuscripts in preparation

I. Published manuscripts

Volchenboum, S.L., **Kristjansdottir,** K., Wolfgeher, D., and Kron, S.J. Rapid validation of Mascot search results via stable isotope labeling, pair picking and deconvolution of fragmentation patterns. *Mol Cell Proteomics.* 2009. **8**, pp. 2011-22.

Kristjansdottir, K., and Kron, S.J. Stable isotope labeling for protein quantitation by mass spectrometry. Review. Current Proteomics. 2010. **7**, pp. 144-155.

II. Manuscripts in preparation

Kristjansdottir, K.*,* Greene, GL., Wu, D. and Kron. S.J. Phosphoprotein profiling of tamoxifen response in MCF-7 cells. *In preparation.*

Volchenboum, S.L., **Kristjansdottir, K.** and Kron, S.J. Identifier, a rapid search engine for high-accuracy stable isotope labeled mass spectrometry data. Modeling protein exclusion . *In preparation.*

Conclusions

I have developed a method for comparison of global phosphoprotein profiles. The methodology involves stable isotope labeling, a phosphoprotein affinity step, 1-D SDS-PAGE and LC-MS/MS. I have performed phosphoprotein profiling of MCF-7 (tamoxifen sensitive) and MCF-7/HER2-18 (tamoxifen resistant) cells as a result of a short (30 minute) tamoxifen treatment. Comparing the results identified 26 proteins that respond to tamoxifen differently in MCF-7 (tamoxifen sensitive) and MCF-7/HER2-18 (tamoxifen resistant) cells. All but three of these proteins are known to be phosphorylated. Several proteins have previously been described as being involved in generation of tamoxifen resistance including FADD and PAK1, showing that phosphoprotein profiling is capable of identifying proteins relevant to tamoxifen resistance.

Examining proteins only found in the phosphoenriched section of one of the cell lines revealed an abundance of proteins involved in apoptosis. One of these proteins, FADD, has previously been shown to result in resistance to tamoxifen when phosphorylation on Ser194 is blocked. We found that FADD was present in phosphoenriched fraction from MCF-7 cells but was not detected in MCF-7/HER2-18 cells. This is not due to changes in proteins amounts since the RT-PCR showed no chance in mRNA and since Western blots showed similar amounts of FADD present in both MCF-7 and MCF-7/HER2-18 cell extracts.

I show in this report that two known phosphorylation sites in XRCC1, Ser447 and Thr453, are detected in the tamoxifen resistant cell line and the levels of these significantly decreased after tamoxifen treatment. Mutations affecting XRCC1 protein levels and activity have previously been associated with increased breast cancer risk. A manuscript describing these results is in preparation.

In collaboration with Sam Volchenboum, Instructor in Pediatrics and the Computational Institute at the University of Chicago, I developed a fast and reliable method for automated validation of Mascot search results from high accuracy mass spectrometry data which was published the paper in Journal of Molecular and Cellular Proteomics. In addition, Jonathan Goya a colleague in the Kron Lab wrote a quantitation module to use for analysis of ^{18}O labeled proteomics data. This manuscript is in preparation.

References

1. Couse JF, Korach KS: **Estrogen receptor null mice: what have we learned and where will they lead us?** *Endocrine Reviews* 1999, **20**(3):358-417.
2. Hunt K: **Breast cancer risk and hormone replacement therapy: a review of the epidemiology**. *Int J Fertil Menopausal Stud* 1994, **39 Suppl 2**:67-74.
3. McGuire WL: **Steroid receptors in human breast cancer**. *Cancer Res* 1978, **38**(11 Pt 2):4289-4291.
4. Shou J, Massarweh S, Osborne CK, Wakeling AE, Ali S, Weiss H, Schiff R: **Mechanisms of tamoxifen resistance: increased estrogen receptor-HER2/neu cross-talk in ER/HER2-positive breast cancer**. *J Natl Cancer Inst* 2004, **96**(12):926-935.
5. **Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group**. *Lancet* 1998, **351**(9114):1451-1467.
6. Fisher B, Anderson S, Tan-Chiu E, Wolmark N, Wickerham DL, Fisher ER, Dimitrov NV, Atkins JN, Abramson N, Merajver S *et al*: **Tamoxifen and chemotherapy for axillary node-negative, estrogen receptor-negative breast cancer: findings from National Surgical Adjuvant Breast and Bowel Project B-23**. *J Clin Oncol* 2001, **19**(4):931-942.
7. Powles TJ: **The case for clinical trials of tamoxifen for prevention of breast cancer**. *Lancet* 1992, **340**(8828):1145-1147.
8. Wolmark N, Dunn BK: **The role of tamoxifen in breast cancer prevention: issues sparked by the NSABP Breast Cancer Prevention Trial (P-1)**. *Ann N Y Acad Sci* 2001, **949**:99-108.
9. Newby JC, Johnston SR, Smith IE, Dowsett M: **Expression of epidermal growth factor receptor and c-erbB2 during the development of tamoxifen resistance in human breast cancer**. *Clin Cancer Res* 1997, **3**(9):1643-1651.
10. Osborne CK, Shou J, Massarweh S, Schiff R: **Crosstalk between estrogen receptor and growth factor receptor pathways as a cause for endocrine therapy resistance in breast cancer**. *Clin Cancer Res* 2005, **11**(2 Pt 2):865s-870s.
11. Massarweh S, Schiff R: **Resistance to endocrine therapy in breast cancer: exploiting estrogen receptor/growth factor signaling crosstalk**. *Endocr Relat Cancer* 2006, **13 Suppl 1**:S15-24.
12. Schiff R, Massarweh SA, Shou J, Bharwani L, Arpino G, Rimawi M, Osborne CK: **Advanced concepts in estrogen receptor biology and breast cancer endocrine resistance: implicated role of growth factor signaling and estrogen receptor coregulators**. *Cancer Chemother Pharmacol* 2005, **56 Suppl 1**:10-20.
13. Ciocca DR, Elledge R: **Molecular markers for predicting response to tamoxifen in breast cancer patients**. *Endocrine* 2000, **13**(1):1-10.
14. Perez-Tenorio G, Stal O: **Activation of AKT/PKB in breast cancer predicts a worse outcome among endocrine treated patients**. *Br J Cancer* 2002, **86**(4):540-545.
15. Gee JM, Robertson JF, Ellis IO, Nicholson RI: **Phosphorylation of ERK1/2 mitogen-activated protein kinase is associated with poor response to anti-hormonal therapy and decreased patient survival in clinical breast cancer**. *Int J Cancer* 2001, **95**(4):247-254.

16. Osborne CK, Schiff R: **Growth factor receptor cross-talk with estrogen receptor as a mechanism for tamoxifen resistance in breast cancer**. *Breast* 2003, **12**(6):362-367.
17. Nicholson RI, Johnston SR: **Endocrine therapy--current benefits and limitations**. *Breast Cancer Res Treat* 2005, **93 Suppl 1**:S3-10.
18. Toi M, Horiguchi K, Bando H, Saji S, Chow LW: **Trastuzumab: updates and future issues**. *Cancer Chemother Pharmacol* 2005, **56 Suppl 1**:94-99.
19. Diaz LK, Sneige N: **Estrogen receptor analysis for breast cancer: current issues and keys to increasing testing accuracy**. *Adv Anat Pathol* 2005, **12**(1):10-19.
20. Benz C, Scott G, Sarup J, Johnson R, Tripathy D, Coronado E, Shepard H, Osborne C: **Estrogen-dependent, tamoxifen-resistant tumorigenic growth of MCF-7 cells transfected with HER2/neu**. *Breast Cancer Res Treat* 1992, **24**(2):85-95.
21. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M *et al*: **Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2**. *N Engl J Med* 2001, **344**(11):783-792.
22. Blagoev B, Ong SE, Kratchmarova I, Mann M: **Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics**. *Nat Biotechnol* 2004, **22**(9):1139-1145.
23. Gronborg M, Kristiansen TZ, Stensballe A, Andersen JS, Ohara O, Mann M, Jensen ON, Pandey A: **A mass spectrometry-based proteomic approach for identification of serine/threonine-phosphorylated proteins by enrichment with phospho-specific antibodies: identification of a novel protein, Frigg, as a protein kinase A substrate**. *Mol Cell Proteomics* 2002, **1**(7):517-527.
24. Kristjansdottir K, Wolfgeher D, Lucius N, Angulo DS, Kron SJ: **Phosphoprotein profiling by PA-GeLC-MS/MS**. *J Proteome Res* 2008, **7**(7):2812-2824.
25. d'Adda di Fagagna F: **Living on a break: cellular senescence as a DNA-damage response**. *Nat Rev Cancer* 2008, **8**(7):512-522.
26. Shay JW, Roninson IB: **Hallmarks of senescence in carcinogenesis and cancer therapy**. *Oncogene* 2004, **23**(16):2919-2933.
27. Suzuki M, Boothman DA: **Stress-induced premature senescence (SIPS)--influence of SIPS on radiotherapy**. *J Radiat Res (Tokyo)* 2008, **49**(2):105-112.
28. Toillon RA, Magne N, Laios I, Castadot P, Kinnaert E, Van Houtte P, Desmedt C, Leclercq G, Lacroix M: **Estrogens decrease gamma-ray-induced senescence and maintain cell cycle progression in breast cancer cells independently of p53**. *Int J Radiat Oncol Biol Phys* 2007, **67**(4):1187-1200.
29. Nicholson RI, McClelland RA, Finlay P, Eaton CL, Gullick WJ, Dixon AR, Robertson JF, Ellis IO, Blamey RW: **Relationship between EGF-R, c-erbB-2 protein expression and Ki67 immunostaining in breast cancer and hormone sensitivity**. *Eur J Cancer* 1993, **29A**(7):1018-1023.
30. Rauch A, Bellew M, Eng J, Fitzgibbon M, Holzman T, Hussey P, Igra M, Maclean B, Lin CW, Detter A *et al*: **Computational Proteomics Analysis System (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments**. *Journal of proteome research* 2006, **5**(1):112-121.
31. Jansen MP, Foekens JA, van Staveren IL, Dirkzwager-Kiel MM, Ritstier K, Look MP, Meijer-van Gelder ME, Sieuwerts AM, Portengen H, Dorssers LC *et al*: **Molecular**

classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J Clin Oncol* 2005, **23**(4):732-740.
32. Barlund M, Monni O, Kononen J, Cornelison R, Torhorst J, Sauter G, Kallioniemi O-P, Kallioniemi A: **Multiple genes at 17q23 undergo amplification and overexpression in breast cancer**. *Cancer Res* 2000, **60**(19):5340-5344.
33. Holm C, Rayala S, Jirstrom K, Stal O, Kumar R, Landberg G: **Association between Pak1 expression and subcellular localization and tamoxifen resistance in breast cancer patients**. *J Natl Cancer Inst* 2006, **98**(10):671-680.
34. Bostner J, Ahnstrom Waltersson M, Fornander T, Skoog L, Nordenskjold B, Stal O: **Amplification of CCND1 and PAK1 as predictors of recurrence and tamoxifen resistance in postmenopausal breast cancer**. *Oncogene* 2007, **26**(49):6997-7005.
35. Hirokawa Y, Arnold M, Nakajima H, Zalcberg J, Maruta H: **Signal therapy of breast cancers by the HDAC inhibitor FK228 that blocks the activation of PAK1 and abrogates the tamoxifen-resistance**. *Cancer Biol Ther* 2005, **4**(9):956-960.
36. Wulfkuhle JD, Speer R, Pierobon M, Laird J, Espina V, Deng J, Mammano E, Yang SX, Swain SM, Nitti D *et al*: **Multiplexed cell signaling analysis of human breast cancer applications for personalized therapy**. *Journal of proteome research* 2008, **7**(4):1508-1517.
37. Matsuyoshi S, Shimada K, Nakamura M, Ishida E, Konishi N: **FADD phosphorylation is critical for cell cycle regulation in breast cancer cells**. *Br J Cancer* 2006, **94**(4):532-539.
38. Volchenboum SL, Kristjansdottir K, Wolfgeher D, Kron SJ: **Rapid validation of Mascot search results via stable isotope labeling, pair picking, and deconvolution of fragmentation patterns**. *Mol Cell Proteomics* 2009.

Appendices

Manuscript: Volchenboum, S.L., **Kristjansdottir,** K., Wolfgeher, D., and Kron, S.J. Rapid validation of Mascot search results via stable isotope labeling, pair picking and deconvolution of fragmentation patterns. *Mol Cell Proteomics.* 2009. **8**, pp. 2011-22.

Manuscript: **Kristjansdottir, K.,** and Kron, S.J. Stable isotope labeling for protein quantitation by mass spectrometry. Current Proteomics. Review. 2010, **7**, pp. 144-155.

Rapid Validation of Mascot Search Results via Stable Isotope Labeling, Pair Picking, and Deconvolution of Fragmentation Patterns*[S]

Samuel L. Volchenboum‡§¶, Kolbrun Kristjansdottir‖, Donald Wolfgeher**, and Stephen J. Kron‖****

Conventional LC-MS/MS data analysis matches each precursor ion and fragmentation pattern to their best fit within databases of theoretical spectra, yielding a peptide identification. Confidence is estimated by a score but can be validated by statistics, false discovery rates, and/or manual validation. A weakness is that each ion is evaluated independently, discarding potentially useful cross-correlations. In a classical approach to *de novo* sequence analysis, mixtures of peptides differing only in a carboxyl-terminal isotopic label yield fragmentation spectra with single, unlabeled *b*-type ions but pairs of isotope-labeled *y*-type ions, facilitating confident assignments. To apply this principle to identification by fragmentation pattern matching, we developed Validator, software that recognizes isotopic peptide pairs and compares their identifications and fragmentation patterns. Testing Validator 1 on a Mascot results file from FT-ICR LC-MS/MS of $^{16}O/^{18}O$-labeled yeast cell lysate peptides yielded 2,775 peptide pairs sharing a common identification but differing in carboxyl-terminal label. Comparing observed *b*- and *y*-ions with the predicted fragmentation pattern improved the threshold Mascot score for 5% false discovery from 36 to 22, significantly increasing both sensitivity and specificity. Validator 2, which identifies pairs by precursor mass difference alone before comparing observed fragmentation with that predicted by Mascot, found 2,021 isotopic pairs, similarly achieving improved sensitivity and specificity. Finally Validator 3, which finds pairs based on mass difference alone and then deconvolutes fragmentation patterns independently of Mascot, found 964 predicted peptides. Validator 3 allowed raw mass spectrometry data to be mined not only to validate Mascot results but also to discover peptides missed by Mascot. Using standard desktop hardware, the Validator 1–3 software processed the 11,536 spectra in the 93-MB Mascot .DAT file in less than 6 min (32 spectra/s), revealing high confidence peptide identifications without regard to Mascot score, far faster than manual or other independent validation methods. ***Molecular & Cellular Proteomics 8:2011–2022, 2009.***

From the Departments of ‡Pediatrics and ‖Molecular Genetics and Cell Biology, §Computation Institute, and **Ludwig Center for Metastasis Research, The University of Chicago, Chicago, Illinois 60637

Received, October 14, 2008, and in revised form, March 16, 2009
Published, MCP Papers in Press, May 11, 2009, DOI 10.1074/mcp.M800472-MCP200

MS/MS combined with informatics analysis is now a uniquely powerful approach for identifying the components of complex protein samples (1–3). Although new technologies have dramatically enhanced the speed, sensitivity, and precision of LC-MS/MS instrumentation (4), data analysis has neither kept pace with nor taken full advantage of these advances. Determining peptide sequences from fragment ion spectra remains a difficult problem, and three main strategies have matured (5). In *de novo* sequencing, the peptide sequence is inferred directly from the fragment ion spectra, and many algorithms have been developed to automate this process, including Lutefisk (6), PepNovo (7), NovoHMM (8), Peptide Identification via Integer linear Optimization (PILOT) (9), and others (10–13). Incomplete fragmentation patterns and low signal to noise (10) make this method difficult to implement as an exclusive means of peptide identification.

The most commonly used method involves comparing experimental MS/MS spectra to theoretical peptide fragmentation patterns derived from protein sequence databases (4) and reporting the best peptide match, which is then propagated forward through the process of determining likely protein components. Several programs are commonly used, including SEQUEST (14, 15), Mascot (16), and X! Tandem (17, 18). What these algorithms share is the determination of a score for a spectrum-peptide match and subsequently a protein identification, and it is the way in which these scores are assigned and interpreted that distinguishes them (19).

The third method for spectrum-peptide matching is a hybrid of *de novo* and database searching (5) in which small lengths of sequence are generated directly from the fragment ion spectra, and these "sequence tags" (20) are used to corroborate spectrum-database matches. Popular implementations of this strategy include DirecTag (21), GutenTag (22), and MultiTag (23). The limitations to this method include the requirement for consecutive fragmentation ions and the reliance on *de novo* algorithms to identify sequence tags.

Database search is highly susceptible to both overreporting false positives (low specificity) and underreporting true positives (low sensitivity). The search engines provide different scoring systems that cannot be directly compared, as the rankings of spectral quality are often based on arbitrary cutoff values. Recent research has focused less on the sequence

matching algorithms themselves but more on the statistics used to evaluate the resulting match scores (24). PeptideProphet was one of the first algorithms developed to evaluate match scores and assign probabilities by evaluating each match with respect to all other peptide assignments. By using machine learning techniques (an expectation-maximization algorithm), PeptideProphet was shown to have high discriminating power for database search results (25). Initially developed for SEQUEST search results, PeptideProphet has been subsequently adapted for use with database search results from Mascot and X! Tandem. These components are combined in Scaffold, a commercial software suite developed by Proteome Software. An alternative approach is to filter the primary data to exclude poor quality MS/MS scans prior to the database search (26), thereby enhancing the likely significance of each reported match.

Using a false discovery rate instead of a false-positive rate is now the standard statistical measure for reporting error rates in data sets with large numbers of features (*e.g.* proteomics or genomics data) (5, 27). Target-decoy searching as an estimate of false discovery rate (FDR)[1] involves first constructing a database of decoy peptides (28, 29), and this strategy is being incorporated into PeptideProphet (30, 31). For each peptide-spectrum match, the target spectrum is queried against a second (decoy) database with characteristics similar to those of the first (*e.g.* a database of reversed or random peptides). Matches to the decoy database are considered false discoveries, and the number of matches above a particular cutoff score threshold is reported. The target-decoy search option is now available in the newest version (version 2.2) of the database search engine Mascot (Matrix Science).

Despite these advances in mass spectrometry, database searching, and statistical approaches to validating matches, the process of analyzing mass spectrometry data remains time-consuming and computer processor-intensive, often requiring several steps and various data transformations (19). To overcome these limitations, we developed a fast and efficient method for peptide identification validation that minimizes the false discovery rate. Our algorithm relies on data from stable isotopic labeling, which is a standard method for quantifying relative protein abundance in complex mixtures (see Ref. 32 and references therein). Carboxyl-terminal labeling methods, including trypsin-catalyzed ^{18}O exchange (33), result in a mixture of pairs of chemically identical but isotopically distinct peptides. The "light" and "heavy" peptides co-elute from HPLC but are readily distinguished by precursor mass (Fig. 1*A*). Each peptide also has an isotopic envelope comprised of isotopologues, molecules that are identical in composition except they can contain any number of isotopes. In the case of trypsin-catalyzed ^{18}O exchange, two ^{18}O atoms are substituted for the two carboxyl-terminal ^{16}O atoms. Comparison of CID fragmentation patterns of carboxyl terminus-labeled light and heavy precursors (or isotopologues) distinguishes *b*-type and *y*-type ions (34, 35). The carboxyl-terminal fragments (*y*-ions) appear as light (^{16}O) and heavy (^{18}O-substituted) forms, but the amino-terminal fragments (*b*-ions) display a single shared mass (Fig. 1, *B–D*).

The technique of using isotopic pairs to enhance peptide identification is not new, and several authors have recognized that isotopic labeling could be used to differentiate carboxyl-terminal from amino-terminal peptide fragments to facilitate peptide sequence analysis (2, 33, 35–38). This method has been productively applied to *de novo* analysis (12, 39–45) and peptide mass fingerprinting (46). In addition, analogous techniques have been applied to the analysis of mixtures of modified and unmodified peptides by probing for peptide mass differences that match known post-translational modifications (47); other groups have used MS/MS spectra information to corroborate these matches and remove noise (48, 49). Finally, isotopic labeling with ^{18}O has been used for manual validation of peptide identifications by observing the predicted mass shift of *y*-ions (50). Nevertheless, this strategy has yet to be harnessed as a means for automated data analysis and peptide search validation.

The goal of this study was to develop a set of software tools designed to provide rapid and automatic validation of peptide assignments by Mascot and to determine the relative benefit of reducing false discovery and the magnitude of loss of *bona fide* identifications. We hypothesized that the characteristic shifting of *y*-type ions between fragmentation spectra of light and heavy precursors might provide a robust check for validity of peptide assignment by database search. Here we demonstrate the feasibility of quickly and efficiently analyzing searched mass spectrometry data, determining within minutes which peptide and protein assignments are likely valid. In its simplest form, Validator 1, identified isotopic pairs in a Mascot results file and improved the 5% FDR cutoff from a Mascot score of 36 to 22, thereby capturing many true identifications that would otherwise have been discarded. A more advanced algorithm, Validator 3, that considers only precursor ion mass, charge, and fragmentation spectral data to identify isotopic pairs independently of any peptide identifications, not only rapidly validated the Mascot results but also discovered peptides that Mascot had failed to match. Our software suite, Validator 1–3, provides new and robust tools for rapid validation of searched LC-MS/MS data obtained in stable isotope experiments, offering improved sensitivity and specificity over database searching alone.

EXPERIMENTAL PROCEDURES

Standardized and Normalized Data Sets—To provide normalized data for our analysis, we prepared a complex soluble protein sample from budding yeast cell lysate. The sample was subjected to prote-

[1] The abbreviations used are: FDR, false discovery rate; ROC, receiver operating characteristic; LTQ, linear trap quadrupole; PME, precursor mass error.

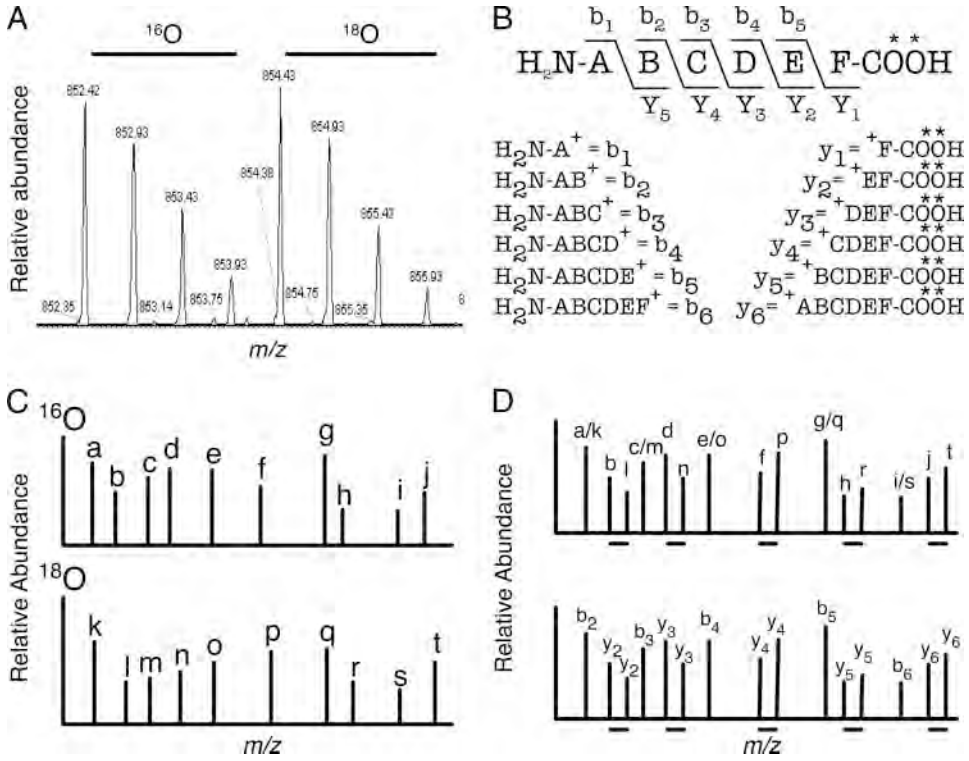

FIG. 1. **Peptide pair identification strategy.** *A*, shown is an example of experimental spectra of a $^{16}O/^{18}O$-peptide pair. Each peptide has an isotopic envelope comprised of three to four different isotopologues containing zero to three molecules of ^{13}C, ^{15}N, or other naturally occurring stable isotopes. The ^{18}O envelope is shifted by about 2.0 Da, reflecting the difference in mass due to the substitution of two ^{18}O atoms. Note that the difference of 2.0 Da is due to the peptide having a 2+ charge state. Peptide pairs with a 1+ charge would be separated by about 4.0 Da. *B*, the *b*-type and *y*-type ions from the collision-induced dissociation of a peptide are shown. Any carboxyl-terminal substitution (as in ^{18}O, indicated by *) will affect the *y*-ions exclusively. *C*, idealized sample MS/MS spectra from the peptide and ions in *B*. The spectra from the ^{16}O- and ^{18}O-peptide forms have similar patterns, although the peak heights may be different. *D*, *top*, the two spectra from *C* are overlaid to demonstrate that the *b*-ions will have a nearly identical mass-to-charge ratio, whereas the *y*-ions will have a shift reflective of the stable isotope substitution. In the example given, peaks "*a*" and "*k*" from *C* are both *b*-ions and therefore overlap, whereas peaks "*b*" and "*l*" are *y*-ions with *l* being shifted due to the substitution of two ^{18}O atoms. Shifted ions are indicated with a *horizontal bar underneath*. By observing which ions overlap and which have shifted, the identities of the *b*- and *y*-ions can be inferred (*D*, *bottom*).

olysis by trypsin. In detail, the proteins were mixed with 6 μl of Rapigest (Waters) and 10 mM tris(2-carboxyethyl)phosphine HCl, denatured at 37 °C for 30 min, alkylated with 10 μl of 50 mM iodoacetamide at room temperature in the dark for 40 min, and digested with 1:50 (w/w) trypsin in 50 mM ammonium bicarbonate, pH 8.9, at 37 °C overnight. The Rapigest was removed by adding 5 μl of 1% TFA. The sample was split and was exchanged in 100% [^{18}O]water or 100% [^{16}O]water using the ^{18}O Proteome Profiler kit (Sigma-Aldrich). MALDI-TOF analysis was used to follow the reaction. Finally this sample was mixed in equal amounts to create a 1:1 ^{16}O:^{18}O reference sample. The resulting peptide mixture was then subjected to reverse phase nanoelectrospray ionization LC-MS/MS on the LTQ-FT instrument (Thermo) using a standard gradient (Zorbax 300SB-C18 column, 150 mm × 75 μm; 0.1% formic acid in water with 5–60% acetonitrile; 0.5%/min gradient). The LTQ-FT instrument was run in positive ion mode at 50,000-ppm resolution MS for ICR. Parent ions were selected for fragmentation by data-dependent analysis using a cycle of one MS scan for ICR (*m/z* 400–2000) and up to five MS/MS scans in the LTQ (*m/z* 50–2000) of the most abundant ions using 120-s dynamic exclusion. A normalized collision energy of 35 was used for low energy CID MS/MS of peptide ions. Under these conditions, a high fraction of the most abundant peptides had both the ^{16}O and ^{18}O monoisotopic species subjected to CID based on our preliminary data. The data set was analyzed by Mascot (version 2.2, Matrix Science) and X! Tandem (version 2007.01.01.1, Global Proteome Machine Organization) to identify peptides and proteins from the MS/MS spectra. Mascot was set up to search the NCBInr_20060910 database (selected for *Saccharomyces cerevisiae*, 11,101 entries) assuming the digestion enzyme trypsin, a fragment ion mass tolerance of 1.0 Da, and a parent ion tolerance of 0.2 Da. Double ^{18}O modification of carboxyl-terminal lysine or arginine, oxidation of methionine, *N*-formylation of the amino terminus, and iodoacetic acid derivative of cysteine were specified as variable modifications. X! Tandem was set to search the scd.fasta.pro database (selected for *S. cerevisiae*, 6,794 entries) also assuming trypsin with a fragment ion mass tolerance of 0.60 Da and a parent ion tolerance of 10.0 ppm. Iodoacetamide derivative of cysteine was specified as a fixed modification. Double ^{18}O modification, deamidation of asparagine and glutamine, oxidation of methionine and tryptophan, sulfone of methionine, tryptophan oxidation to formyl, and acetylation of lysine and the amino terminus were specified as variable modifications. Scaffold (version Scaffold-01_06_00, Proteome Software) was used to validate MS/MS-based peptide and protein identifications. Peptide identifications are accepted if they can be established at greater than 90.0% probability as specified by the PeptideProphet algorithm (51). Protein identifications are accepted at greater than 95.0% probability and contain at least one identified peptide with probabilities assigned by the ProteinProphet algorithm. Proteins that contain similar peptides

TABLE I
Validator data

For each version of Validator, the number of pairs, queries, and queries with peptides is shown. In addition, data are displayed after filtering the raw Mascot data for only those peptides with scores greater than 35. The precursor mass error range corresponds to the *dotted* ("all") and solid (">35") *lines* in Fig. 3. NA, not applicable.

Version	Raw	Raw >35	1	2	2e	3	3e
Pairs identified	NA	NA	2,775	3,209	NA	3,779	2,021
Mascot queries	20,759	2,308	2,345	3,185	1,782	3,615	2,310
Queries with peptides	17,200	2,308	2,345	3,177	1,782	3,545	2,289
PME range (±) with 95%: all	0.193	0.024	0.022	0.134	0.042	0.142	0.129
PME range (±) with 95%: >35	0.024	0.024	0.017	0.011	0.011	0.011	0.013
Unique peptides	13,158	580	398	1,564	481	1,881	964
Unique proteins	5,962	186	125	1,150	234	1,391	696
Score at FDR 5%	36	36	22	36	29	37	37
Score at FDR 2%	42	42	32	41	34	43	43
Percentage of queries with Mascot score >35	13.4	100	78.0	46.6	75.2	42.1	57.1

and cannot be differentiated based on MS/MS analysis alone are grouped to satisfy the principles of parsimony.

Software Development—All software analysis was performed on searched Mascot data (*e.g.* ".DAT files"). Custom software was written in Python 2.6. Statistical analysis was performed using both Python scripting as well as Microsoft Excel. Charts and graphs were generated using both Python's Matplotlib library (SourceForge, Inc.) and GraphPad Prism. Software was run on standard desktop and laptop computers running both Windows XP (service pack 3) and Macintosh OS 10.5. Details about software development and implementation are included under "Results."

RESULTS

The aim of this study is to describe a fast and efficient means for validating peptide identifications obtained by searching ^{18}O-labeled MS/MS data with Mascot. Our approach is to mine the Mascot .DAT file to extract information not utilized by Mascot but potentially useful for automated validation. For the purposes of this study, we refer to a "query" as any precursor ion and its associated fragmentation ions, regardless of whether Mascot assigned a match, and to a "peptide" as any query to which Mascot assigned a match, regardless of Mascot score and without external validation. For each query, up to 10 possible peptides are assigned by Mascot, each with a probability score. For this study, we examined all query-peptide identifications as well as only the top scoring match suggested by Mascot. Using a $^{16}O/^{18}O$-labeled data set from yeast cell lysate, analysis of the Mascot .DAT file revealed 20,759 queries and 17,200 peptide identifications, corresponding to 13,158 unique peptides and 5,962 unique proteins, using only the top suggested Mascot peptide identification (Table I). The FDR of 5% was achieved at a threshold Mascot peptide score of 36, and 2% was achieved at a cutoff score of 42.

The majority of peptides have low Mascot scores (Fig. 2*A*). As expected, peptides with the highest Mascot scores tend to have a low precursor mass error (PME) (Fig. 3*A*). In fact, the search results represent two populations: peptides with high Mascot score/low PME and peptides with low Mascot score/high PME. A plot of the Mascot score *versus* the variance of the PME for all peptide matches above that score illustrates a steep fall in the variance, plateauing close to a Mascot score of 35 (supplemental Fig. 1), providing an approximate cutoff threshold separating the two populations. Of the 17,200 peptides identified by Mascot, 2,308 have scores greater than 35. The width of precursor mass error range that encompasses 95% of these peptides with high Mascot scores is 0.048 Da, whereas the interval that covers 95% of all peptides is 0.386 Da (Fig. 3).

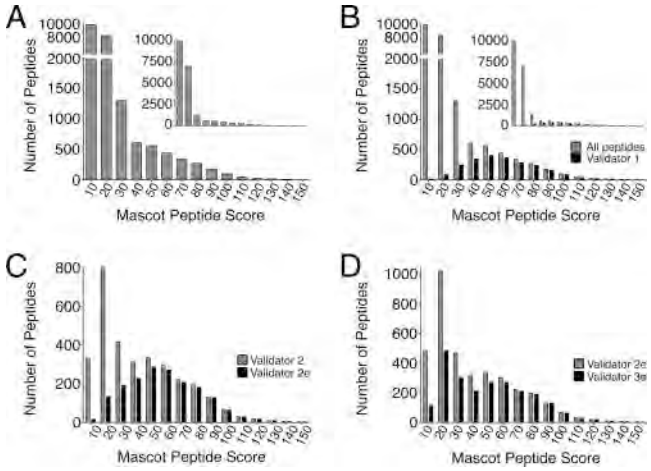

FIG. 2. **Distribution of Mascot scores.** *A*, the raw Mascot data file was parsed, and the number of peptides in each score group was tallied. The vast majority of scores were less than 30. Note that the *y axis* has a break at 2,000. See the *inset* for the full-scale graph with identical *x axis* but no break in the *y axis*. *B*, Validator 1 finds $^{16}O/^{18}O$ pairs in the searched Mascot data file. The distribution of Validator 1-derived peptide scores (*black*) is seen against the raw distribution (*gray*) from *A*. Again, note the broken *y axis* and the *inset* showing the full *y axis* scale. At the low end of the scores, Validator 1 rejects most of the peptides while retaining most of the high scoring peptides. *C*, the Validator 2e-identified peptides with fragment ion tallies greater than 10 (*black*) are shown compared with the Validator 2 results (*gray*). At low scores, Validator 2e rejects most low scoring peptides while retaining most peptides with high Mascot scores. *D*, Validator 3e (*black*) performs similarly to Validator 2e (*gray*) despite not utilizing any Mascot search information.

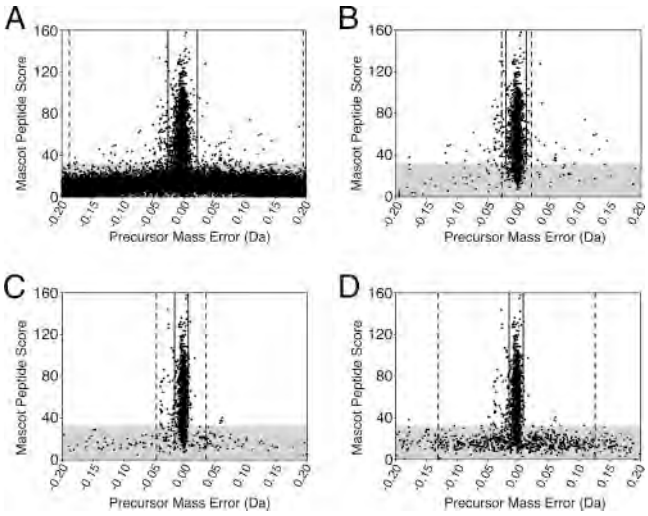

FIG. 3. **Precursor mass error *versus* Mascot score.** Low Mascot peptide scores, as defined as a score less than 35, are shown in the *shaded gray area*. *A*, the raw data are separated into two distinct zones: the high Mascot score peptides, most with low precursor mass error, and the low Mascot score peptides, most with high precursor mass error. As the Mascot score increases from 0 to 35, the variance of the precursor mass errors of all peptide matches above this score falls dramatically (see also supplemental Fig. 1). We determined cut-offs for precursor mass error that would encompass 95% of all peptides (*dashed lines*) and 95% of peptides with Mascot peptide scores over 35 (*solid lines*). *B*, Validator 1 successfully removes most of the peptides with low Mascot peptide scores. Note the more narrow 95% range for all peptides (*dashed lines*) compared with *A* as well as the much tighter 95% interval for peptides with Mascot peptide scores greater than 35 (*solid lines*). *C*, Validator 2e-identified peptides with a fragment ion tally of 10 or more are shown. Note that although the interval encompassing 95% of the peptides (*dashed lines*) is wider than for Validator 1 it is much narrower than for the raw data. In addition, the 95% interval for peptides with Mascot peptide scores greater than 35 (*solid lines*) is narrower than for Validator 1-identified peptides. *D*, Validator 3e-identified peptides with a fragment ion tally of at least 10 are shown. Again the intervals encompassing 95% of the peptides (*dashed lines*) and 95% of peptides with Mascot scores greater than 35 (*solid lines*) are shown.

Validator 1—As a proof of concept, we first sought to find all $^{16}O/^{18}O$ pairs in the Mascot summary file (".DAT file"). Here a $^{16}O/^{18}O$ pair refers to a peptide sequence identified in two distinct isotopic forms in the same Mascot file as an unlabeled ^{16}O-peptide and as a peptide containing two ^{18}O atoms. The ^{18}O form of each peptide is 4.008491 Da heavier than its unlabeled ^{16}O form (Unimod). Our first program, Validator 1, is designed to utilize the peptide identifications made by Mascot. Validator 1 first iterates through all queries looking for identical top scoring peptides found in both ^{16}O and ^{18}O forms (a "$^{16}O/^{18}O$ pair"). As the ^{16}O and ^{18}O forms are expected to co-elute from reverse phase columns, we added a constraint that the MS/MS scans of the two peptides must occur within 200 scan units (~2.25 min) of each other. With these criteria, Validator 1 identified 2,775 pairs representing 2,345 unique matched queries with peptides. These peptides represented 398 unique peptides and 125 unique proteins (Table I). This analysis required ~10 s of calculation on a laptop computer. The precursor mass range width that encloses 95% of the peptides with Mascot scores greater than 35 was 0.034 Da, whereas the width of the range that encompasses 95% of all peptides decreased by 89% compared with Mascot alone, to 0.044 Da (Fig. 3, *A versus B*).

There were 223 unique peptides with Mascot scores over 35 that Validator 1 failed to discover as a member of a $^{16}O/^{18}O$ pair. Manual examination of the raw spectra for 10 of the highest scoring of these peptides revealed three scenarios. For six peptides, the ^{16}O form was fragmented and yielded a high Mascot score, but the ^{18}O form was not selected for MS/MS. In one case, the ^{18}O form subjected to MS/MS was an isotopologue not accounted for by the Mascot search and thus was not correctly identified. In three cases, a candidate pair was flagged by Validator 1, but the data turned out to correspond to two peaks within the isotopic envelope of a single peptide.

On the other hand, Validator 1 did not reject all low scoring peptides, particularly where the Mascot identifications yielded low precursor mass errors. As seen in Fig. 3*B*, these peptides represent a "comet tail" in the data, stretching all the way down to Mascot scores as low as 10. A closer inspection of these peptides (data not shown) reveals that most were also found in other queries with high Mascot scores. Nevertheless, of the low scoring peptides found by Validator 1, there were 21 proteins represented that would not be identified if only high Mascot scoring peptides were being retained.

Therefore, Validator 1 was able to rapidly identify $^{16}O/^{18}O$ pairs within searched Mascot data. Using $^{16}O/^{18}O$ pairs as a criterion rather than a simple Mascot threshold retained most high scoring peptides and rejected most low scoring peptides but also rescued several low scoring but likely correct identifications.

Validator 2—Validator 1 relies on Mascot to identify both the ^{16}O- and ^{18}O-labeled peptides. We reasoned that additional $^{16}O/^{18}O$ pairs might be found in the Mascot .DAT file by searching for pairs of queries where the precursor masses were separated by a difference of 4.008491 Da without regard to any features of the MS/MS data or whether Mascot had assigned the same, different, or even any identifications. Thus, the Validator program was modified to start with a query identified as a ^{16}O- or ^{18}O-peptide and search the Mascot .DAT file for queries within a range of 200 scan units (2.25 min) with a precursor mass difference of 4.008491 Da and with a mass error limit of 3 ppm. Using these criteria, Validator 2 found 3,209 pairs representing 1,564 unique peptides and 1,150 unique proteins.

The most significant distinction between Validator 1 and 2 was the retention of considerably more low scoring peptides. Notably, of the 3,177 peptides retained by Validator 2, 1,696 had Mascot scores below 35, and many also displayed a high mass error, suggesting a low likelihood of correct identification. These results raised the question of whether using ad-

ditional criteria based on the MS/MS data embedded in the Mascot data file might help reveal potentially correct peptide matches with low Mascot peptide scores while filtering out incorrect identifications.

Validator 2e—Given that fragmentation spectra are available for each member of a candidate $^{16}O/^{18}O$-peptide pair identified by Validator 1 or 2, we hypothesized that these data could be mined to distinguish false identifications. As noted above, comparing the MS/MS fragmentation of the light and heavy forms will reveal identical sets of *b*-ions but distinct *y*-ions with pairs of fragments shifted by 4.008491 Da, reflecting the exchange of two ^{18}O atoms for ^{16}O at the carboxyl terminus (Fig. 1). We therefore extended our program, dubbed Validator 2e, to take advantage of the embedded carboxyl-terminal labeling information to distinguish the *b*-type and *y*-type ions, facilitating peptide validation.

As a first step, we confirmed that the MS/MS ions in each query correspond with a theoretical fragmentation table based on the sequence of the peptide match provided by Mascot. For each peptide identification in the Mascot data file, we calculated the fragmentation table and counted the number of observed ions that fell within a window of 2000 ppm from a predicted *b*- or *y*-ion. As expected, there is a positive correlation between the number of *b*- and *y*-ion matches and Mascot peptide score ($r = 0.596$, $p < 0.0001$; supplemental Fig. 2A). To validate Mascot identifications for $^{16}O/^{18}O$ pairs, we tested whether the following held true: when pairs of ions matched predicted *b*-type ions, they should be identical (non-shifting), whereas those matching *y*-ions should differ by 4.008491 Da (shifting). The number of matching pairs of non-shifting *b*-ions and shifting *y*-ions were thus tallied to generate a "fragment ion tally." We hypothesized that a high fragment ion tally would characterize a correct peptide identification for a query member of a $^{16}O/^{18}O$ pair.

For each pair identified by Validator 2, we calculated the fragment ion tally for each query member based on comparison with predicted fragmentation tables for the highest scoring peptide match provided by Mascot. Fragment ion tally correlates with a high Mascot peptide score ($r = 0.639$, $p < 0.0001$; supplemental Fig. 2B) with a fragment ion tally of 10 corresponding to a Mascot score of 35. We therefore filtered the list generated by Validator 2 to retain only pairs that yielded a fragment ion tally of at least 10 with at least two matching shifting (*y*-type) ions. The requirement of two *y*-ion (shifting) matches will reject pairs of ions derived from the same isotopic envelope that are predicted to yield many matching *b*-ions but no matching *y*-ions. Calculating fragment ion tallies for the 3,209 pairs of queries found by Validator 2 yielded 1,782 queries with counts greater than or equal to 10 (Table I). These queries represent 481 unique peptides and 234 proteins. Notably, of the query-peptide matches with fragment ion tallies of 10 or greater, only 442 (24.8%) had Mascot scores less than 35. Compared with Validator 2, Validator 2e eliminates many of the low scoring/high mass error peptides but retains most of the high scoring/low mass error peptides (Fig. 2*C*). Limiting the plot to peptides evaluated with Validator 2e that yield a fragment ion tally of 10 or greater, 95% of high scoring peptides fell within a precursor mass error range of 0.022 Da *versus* a range of 0.084 Da for all peptides (Fig. 3*C*). Compared with Validator 1, Validator 2e found 219 queries, 163 peptides, and 135 proteins not found by Validator 1 (supplemental Table 1).

Validator 3/3e—As a next logical step, we sought to find candidate pairs based solely on their mass difference and ion lists from raw data without regard to any peptide sequence information provided by Mascot in the .DAT file. Validator 3 identifies pairs much like Validator 2 except for not requiring that one member of the pair be a Mascot-identified ^{16}O- or ^{18}O-peptide. The program iterates through all queries and searches for another query with the predicted 4.008491-Da mass difference, allowing an error of 3 ppm. From the reference data set, the program identified 3,779 pairs, representing 3,615 unique queries, of which 3,545 have Mascot-assigned peptide identifications. Examination of the data revealed that some Validator 1 pairs remained unidentified, as their difference in precursor mass lies outside the 3-ppm tolerance limit imposed by Validator 3 (data not shown). Validator 3 found 1,875 queries, 1,540 peptides, and 1,279 proteins not found by Validator 1 (supplemental Table 1).

As with Validator 2e, we extended Validator 3 to 3e by utilizing the expectation of non-shifting *b*-ions and shifting *y*-ions to perform an internal validation of the proposed pairs, without relying on the peptide identification(s) provided by Mascot. Therefore Validator 3 was modified to find pairs of shifting and non-shifting fragment ions for each pair based on comparing the two lists of MS/MS ions and finding non-shifting *b*-ions and shifting *y*-ions within a mass tolerance of 2,000 ppm. To decrease the influence of noise, only fragment ions with a peak height of at least 0.5% of the intensity of the strongest ion were evaluated. To be considered a shifting or non-shifting pair, the difference in intensity between the heavy and light forms of the candidate could be no more than 25%. Again a fragment ion tally was determined from the number of pairs of candidate *b*- (non-shifting) and *y* (shifting)-ions while requiring at least two *y*-ions. To validate the scoring scheme, the fragment ion tally and Mascot peptide scores were compared, and as with Validator 2e, we found a significant positive correlation ($r = 0.395$, $p < 0.0001$; supplemental Fig. 2C).

Because two complete sets of MS/MS ions are being compared without regard to a predicted fragmentation pattern, we expected to identify more pairs with higher fragment ion tallies. To facilitate comparison with Validator 2e, we filtered based on a fragment ion tally cutoff of 10, yielding 2,310 queries (Table I). These correspond to 964 peptides and 696 proteins identified. As expected, Validator 3e was less selective than Validator 2e in rejecting low scoring peptides (Fig. 2*D*) while retaining a higher proportion of high mass error

TABLE II
Scaffold comparison

Results are shown comparing the performance of Validator versions 1–3 with the peptide and protein output from the commercial software package Scaffold. In addition, data are displayed after filtering the raw Mascot data for only those peptides with scores greater than 35. The Scaffold filtering criteria were to include only peptides with a 90% confidence, proteins with a 95% confidence, and only those for which there were at least two unique peptides identified. For instance, using only the top peptide match from Mascot for each query, Validator 1 captured 69.5% of the peptides and 91.9% of the proteins as identified by Scaffold. Also shown are results when using all possible peptide and protein guesses by Mascot. ID'd, identified.

Version	Raw	Raw >35	1	2	2e	3	3e
Top Mascot query match							
Percentage of Scaffold peptides ID'd	99.6	99.4	69.5	66.1	62.6	67.1	59.1
Percentage of Scaffold proteins ID'd	100	100	91.9	93.0	84.9	94.2	88.4
Percentage of peptides ID'd not in Scaffold	96.4	18.8	18.6	80.4	39.7	83.4	71.7
Percentage of proteins ID'd not in Scaffold	97.5	56.8	47.6	90.2	64.8	91.6	84.7
All Mascot query matches							
Percentage of Scaffold peptides ID'd	100	99.8	71.1	68.9	64.4	69.9	60.7
Percentage of Scaffold proteins ID'd	100	100	97.7	98.8	95.3	98.8	96.5
Percentage of peptides ID'd not in Scaffold	99.5	96.7	95.9	98.5	97.4	98.6	98.1
Percentage of proteins ID'd not in Scaffold	98.2	97.6	96.9	97.9	97.5	97.9	97.7

peptides (Fig. 3*D*). The precursor mass error range containing 95% of peptides with scores greater than 35 was quite similar to that of Validator 2e, 0.026 *versus* 0.022 Da, but considerably wider for all peptides, 0.258 *versus* 0.084 Da. These data show that a strategy agnostic to Mascot-specific peptide information can be used to identify peptides highly likely to represent *bona fide* $^{16}O/^{18}O$ pairs, providing independent validation for Mascot identifications.

Comparison with Scaffold—The commercial proteomics software suite Scaffold (Proteome Software) uses the PeptideProphet algorithm (25) to generate lists of peptides and proteins with an associated probability. Many groups use Scaffold for downstream data analysis, and we feel that it is important to compare the performance of our software with that of this commonly used analysis tool. Using the same Mascot .DAT file, the data were analyzed in Scaffold using probability cutoffs for peptides and proteins of 90 and 95%, respectively. The list of proteins meeting these criteria along with the constituent peptides was compared with the peptide and protein lists generated by Validator versions 1–3e (Table II). Using the top scoring Mascot peptide identifications only, Validator 1 found 69.5% of the peptides and 91.9% of the proteins found by Scaffold. The performance of Validator 2e was similar, identifying 62.6 and 84.9% of the peptides and proteins, respectively. Validator 3e found 59.1% of the peptides and 88.4% of the proteins found by Scaffold. The seven proteins identified by Scaffold but not identified by Validator 1 were examined. Four proteins had peptide pairs with the MS mass difference outside of the Validator 3e tolerance of 3 ppm. One protein had a fragment ion tally below the cutoff limit of 10. Two proteins were identified solely from ^{16}O-peptides with no ^{18}O partner and would thus not be identified by any form of the Validator software.

Corroboration of Validator 1-identified Peptide Pairs—Returning to the $^{16}O/^{18}O$ pairs identified by Validator 1, we sought to corroborate the pairs by analysis of shifting and non-shifting fragment ions. The Validator 3e program was extended to analyze all Validator 1-identified pairs, first by finding all shifting and non-shifting ions between the two MS/MS ion lists. Then the list of matches was compared with the predicted fragmentation table for the Mascot-identified peptide to calculate a fragment ion tally. To determine the significance of each potential match, the following algorithm was used: for each potential peptide pair, we randomly permuted the peptide sequence 30 times, each time computing the fragmentation table for the random peptide and determining a fragment ion tally. Based on the distribution of fragment ion tallies for the randomly permuted peptides, a 95% confidence interval was determined. Using a criterion that the fragment ion tally for the Mascot-identified peptide must fall outside this range, the fragment ion tallies for 2,626 (94.6%) of the 2,775 Validator 1-identified peptides were found to be significant. In other words, using internal pair validation based on matching shifting and non-shifting MS/MS ions, we were able to corroborate almost every $^{16}O/^{18}O$ pair found by Validator 1. This is highly significant as it both demonstrates the strength of using $^{16}O/^{18}O$ pair finding as a route to high confidence peptides and validates our method of peptide validation by matching MS/MS ions.

Statistical Analyses—We next sought to analyze our results by applying a conventional validation method of false discovery rate determination and receiver operating characteristic (ROC) curve plotting. Whenever a protein sequence from the target database is tested, a random sequence of equal length and similar amino composition is generated and tested (Matrix Science and Refs. 29 and 52). Any matches to the decoy database are assumed to be false positives, and this approach assumes that matches to the decoy peptides have the same distribution as false-positive matches to the original target data (5). For calculation of FDR at a given threshold

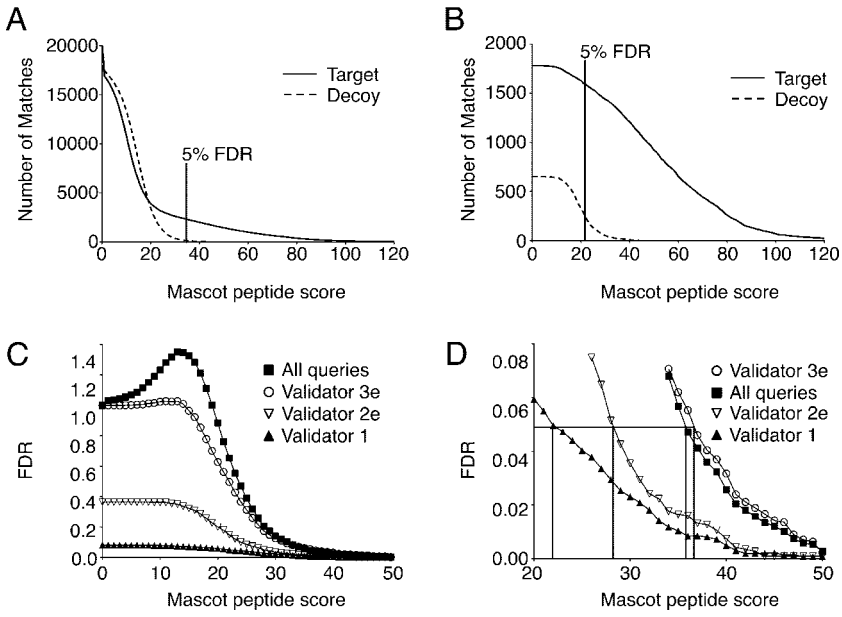

FIG. 4. **Analysis of FDRs.** *A*, number of Mascot peptide-spectrum matches for target (*solid*) and decoy data (*dotted*). The total number of matches with peptide scores over the given Mascot cutoff score is shown, and the score threshold for an FDR of 5% is indicated. *B*, number of Validator 2e matches for target data (*solid*) and decoy data (*dotted*). Note the different *y axis* scale compared with *A*. *C* and *D*, false discovery rate for raw Mascot and data filtered by Validator versions 1, 2e, and 3e. False discovery rate is the number of decoy peptides divided by the number of target peptides with scores exceeding a given threshold. In *D*, the *black lines* mark the Mascot peptide score cutoffs to achieve an FDR of 5% for Mascot (35.6) and Validator 1 (22), 2e (29), and 3e (37).

score, we used the method described by Käll *et al.* (27, 29) of dividing the number of decoy peptides identified (with scores over the threshold) by the number of target peptides identified (with scores over the threshold score). In general, the identified decoy peptides have low Mascot peptide scores and high precursor mass errors (supplemental Fig. 3). Searching the data set with Mascot against the reference proteomes of 17,200 target peptides and 17,687 decoy peptides yielded an FDR of 5% at a Mascot peptide score of 36 (Fig. 4*A*). At this cutoff score, Mascot retains 2,250 target peptides and 106 decoy peptides. We were interested in comparing the features of decoy peptides as an independent means of estimating the ability of Validator to decrease FDR. We therefore applied this test to analyze the filtering ability of Validator versions 1–3 (Table I). As an example, recall that Validator 2e identifies pairs by first finding a pair member that Mascot has identified as having either a carboxyl-terminal ^{16}O or ^{18}O and then finding the other pair member by searching for a peptide with the appropriate difference in *m/z*. Using this Mascot-identified peptide for each pair member, the program identifies the *b*- and *y*-ions from the list of MS/MS ions. This list is searched against the list of MS/MS ions from the isotopic partner to determine the number of non-shifting (*b*-type) and shifting (*y*-type) ions, and the sum of these is the fragment ion tally. Peptide-spectrum matches with a fragment ion tally of 10 or greater are retained. Validator 2e retains 1,782 target but only 650 decoy peptides. The majority of decoy peptides have a low Mascot score so that an FDR of 5% is achieved at a cutoff score of 29 (Fig. 4*B*). At that score, the algorithm retains 1,457 target peptides and 62 decoy peptides.

Receiver operating characteristic curves are a useful way to visualize the relationship between the sensitivity and specificity of a test. We used ROC analysis to probe the relationship between sensitivity and specificity for Mascot peptide scores over all data, prefiltered data, and Validator-filtered data. For a typical mass spectrometry experiment, a true ROC curve cannot be plotted because the true-positive rate is unknown. Typically the search results from the target and decoy data sets are used to approximate the sensitivity and specificity of the search engine filter (Matrix Science). Sensitivity is approximated by the ratio of the number of queries with peptide scores above a given value to the total number of queries. Likewise specificity is approximated by the ratio of the number of decoy queries with assigned peptides above a given score to the total number of decoy peptides. ROC analysis of the full set of Mascot-searched data demonstrates poor sensitivity and specificity throughout most of the range of score thresholds (Fig. 5*A*, *stars*). It is only at a very low threshold score that the sensitivity approaches 100% (capturing all correct identifications) while the specificity is close to zero (capturing all incorrect identifications). As expected, restricting the ROC analysis to peptides with Mascot scores above 10 or above 35 (Fig. 5*A*, *solid* and *open squares*) improves sensitivity and specificity. When the Validator 1 filtering algorithm is applied to the data (Fig. 5*A*, *triangles*), the ROC curve demonstrates a stronger relationship between sensitivity and specificity with a sensitivity of 80% and specificity of 89% at a threshold score of 35 (Fig. 5*A*, *arrow*). The performance of Validator versions 2, 2e, and 3e are similarly compared in Fig. 5*B*. Note that Validator 2e has the best ROC curve with a sensitivity of 80% and a specificity of 94% at a Mascot peptide score threshold of 32 (Fig. 5*B*, *arrow*).

Corroboration of Validator 3-identified Peptide Pairs—A schema for corroboration of Validator 3-identified peptide pairs is shown in Fig. 6. For the pairs identified by Validator 3e, we utilized the Mascot information, where available, to determine the significance of the match. If the Mascot identification was the same for both members of the pair, we determined

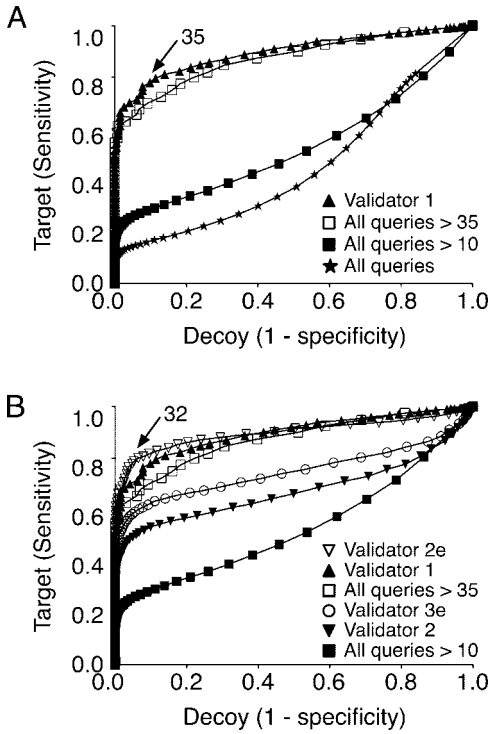

FIG. 5. **ROC curves.** For a given threshold Mascot peptide score, the sensitivity is the ratio of the number of identifications with scores greater than the cutoff score to the total number of queries, whereas the specificity is the ratio of the number of decoy peptide identifications over the cutoff score to the total number of decoy peptide identifications. *A*, ROC curves for Mascot-searched data and Validator 1-filtered peptides. Validator 1 (*triangles*) outperforms a simple score cutoff of 35 (*open boxes*). *B*, ROC curves for Validator versions 1–3. Both Validator 1 and 2e outperform using a simple Mascot score cutoff of 35 (*open boxes*).

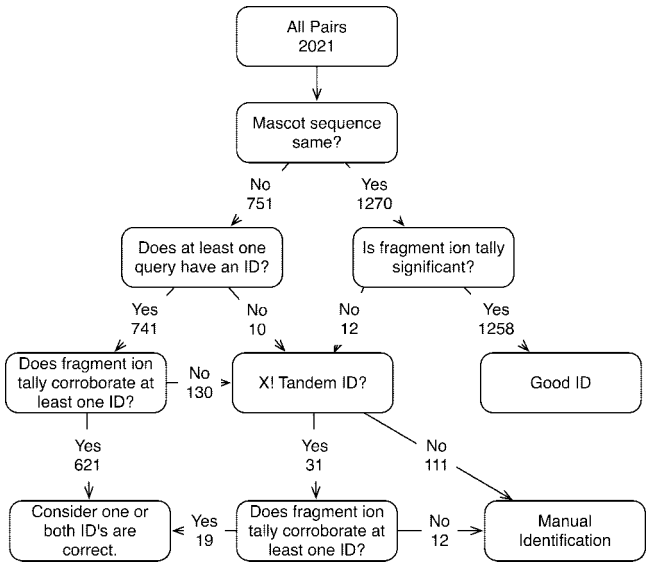

FIG. 6. **Schema for corroborating Validator 3e-identified peptide pairs.** The tallies reflect the results for the test data set. If the Mascot identification (*ID*) was the same, the shifting and non-shifting ions were matched against the fragmentation table. 1,258 of 1,270 pairs were corroborated this way. Of the remaining pairs, if at least one had a Mascot identification, the shifting and non-shifting ions were compared with the theoretical fragmentation table, and if one or both had a valid fragment ion tally, it was assumed correct. This was true for 621 pairs. Of the remaining pairs, a search was performed using X! Tandem, an alternate search engine, and if a peptide was identified, the corroboration was repeated. For 31 peptides, an identification was made using X! Tandem, and for 19 of these, the match was corroborated with the identified ions. For the remaining pairs (133 in this case), a manual review will need to be performed to determine the identity of the peptide and the validity of the match.

the significance of the match using the corroboration strategy of determining fragment ion tallies after randomization of the candidate peptide. Of the 1,270 pairs where the peptide identifications were the same, the score was found to be significant in 1,258 pairs. For the 741 cases where the Mascot identifications were to different sequences, or only one member of a pair had an identification, the same technique was applied to determine the significance. In 621 cases, the corroboration score was significant for at least one matched peptide. For the 130 pairs where there was no corroboration or where neither peptide had a Mascot identification, 31 could be identified using X! Tandem. Of these, we were able to corroborate 19 using the randomization strategy. This left only 133 pairs that passed the fragment ion tally threshold of 10 but lacked any peptide identification to validate. Overall we were able to corroborate 1,898 of 2,021 Validator 3e pairs (93.9%).

Performance—All versions of Validator are written in Python version 2.6 running on desktop and laptop hardware. Versions were tested both in Windows XP and Mac OS X environments. Our reference Mascot .DAT data file is 92.8 MB and 1.24 million lines, consisting of 11,536 scans, 20,759 queries, and their analysis. On standard hardware (*e.g.* Intel Core-2 Duo processors with 2–4 GB of RAM), Validator versions 1–3 run in sequence in less than 6 min (~32 spectra/s), including a complete parsing of the .DAT file, pair finding, and corroboration and full FDR analysis. Validator 1 by itself runs from start to finish in 70 s. Most of this time is spent building the query dictionaries, and once loaded, Validator 1 is able to find all $^{16}O/^{18}O$ pairs in about 10 s, including decoy search and false discovery rate determination. This corresponds to processing >1,000 spectra/s. Once optimized and compiled, it is expected that Validator should be able to run several times faster. To facilitate further development, software will be available freely both as stand alone code as well as a Web-based tool (www.msvalidator.org).

DISCUSSION

We have developed Validator, a novel proteomics database search validation software that provides a direct and independent means to validate peptide identifications provided by Mascot analysis of tandem mass spectrometry data. Our algorithm is based on LC-MS/MS analysis of a mixture of carboxyl-terminal stable isotope-labeled and non-labeled peptides, a common sample in quantitative mass spectrom-

etry (32, 53–57). We exploit the characteristic fragmentation of isotopically labeled peptides to enhance their identification, a well established principle that goes back to the period preceding the modern era of ESI and LC-MS/MS (36, 37) and has since been applied effectively by a number of investigators (*e.g.* Refs. 2, 5, 12, 14, 33, 35, 38–48, and 50). Where both the light (unlabeled) and heavy (labeled) forms of a peptide are selected for fragmentation, the resulting spectra can be compared, thereby distinguishing pairs of non-shifting *b*-ions from pairs of *y*-ions that display a shift determined by the isotopic label. These data are then used to test the validity of Mascot peptide identifications, comparing observed with predicted fragmentation patterns. We found that this approach allows rapid and efficient automated filtering of Mascot analysis of LC-MS/MS data to improve both the sensitivity and specificity of peptide identification while salvaging potentially useful low scoring peptides not captured by conventional validation strategies.

Our naive, first approach was to rapidly identify all Mascot-derived $^{16}O/^{18}O$ pairs from a Mascot .DAT file where both peptides received the same identification. Our data show that a majority of the highest scoring peptides are validated by this simple strategy, and this method was not only able to find 91% of the proteins identified by the commercial analysis package Scaffold but also to capture peptides where the Mascot scores would have fallen below any standard significance threshold. This analysis takes less than 10 s and results in a list of very high confidence peptide and protein identifications. The surprising performance of this simple approach probably reflects the high bar required for Mascot to independently match each of the fragmentation spectra to the ^{16}O and ^{18}O forms of the same peptide, even when the resulting scores fall below normal significance thresholds. In turn, this single criterion efficiently rejects most false identifications as from decoy data.

Validator 2 relaxes the requirement for Mascot to make the same identification for both spectra in a pair and simply seeks a partner for each ^{16}O- or ^{18}O-labeled peptide based on the expected difference in precursor mass. We have shown that this is also a fast and reliable way of identifying pairs, and we found many $^{16}O/^{18}O$-labeled potential matches not identified by Validator 1. With Validator 2e, we extracted the *b*-type (non-shifting) and *y*-type (shifting) fragment ions from the MS/MS spectra of each pair and then compared these data with the theoretical peptide fragmentation table calculated from the Mascot peptide identifications. Validator 2e confirmed both low and high scoring Mascot identifications but also rejected many others, including nearly all high scoring matches to the decoy database. Thus, Validator 2e was able to achieve an FDR of 5% at a score of 29 *versus* 36 for Mascot alone. These data suggest that for any arbitrary level of significance running Validator can significantly increase confidence in peptide identifications independently of the Mascot score.

To develop a validation scheme agnostic to Mascot-derived information, we reasoned that peptide pairs could be found based only on the difference in precursor mass. Validator 3 was able to quickly find all Validator 2-identified pairs as well as many others. Here, even though in many pairs neither the light nor heavy forms were matched by Mascot, we again wanted to corroborate the peptides by matching shifting and non-shifting ions. By comparing the two MS/MS ion series directly, shifting and non-shifting ions were rapidly identified by Validator 3e, and we were able to confirm the majority of high Mascot scoring peptides by tallying the number of shifting and non-shifting ions and again efficiently reject Mascot decoy matches. In addition, Validator 3e validated many pairs that had received low Mascot scores and even determined fragmentation patterns for pairs of queries for which Mascot had made no assignments at all.

Using this fragment ion matching scheme, we were able to corroborate most of the 2,775 pairs found by Validator 1. To study Validator 3-identified peptides, we applied a more complicated but systematic approach and corroborated 94% of peptide pairs by combining multiple analysis methods including X! Tandem and manual validation. These results demonstrate that we can quickly ($<$5 min) parse a Mascot results file, returning a list of high confidence peptide pairs, many of which would be missed using conventional score cutoff techniques.

Because our software is designed to analyze data from samples that are a mixture of peptides labeled at the carboxyl terminus with either ^{16}O or ^{18}O, there is some concern that MS analysis of the mixture will result in fewer protein identifications than for an unlabeled sample due to an increase in fragmentation of "redundant" isotopologues at the expense of other peptides. Indeed when we analyzed ^{16}O and ^{18}O samples separately, we found that Mascot identified about 30% more peptides in either singly labeled sample than when the MS was performed on the 1:1 mixture. Thus, we modified Validator to allow for separate ^{16}O and ^{18}O fractions to be combined and analyzed as a single data set, and as expected, analysis of the combined fractions rescues the lost identifications (data not shown). Whether analyzed separately (requiring more MS time) or together (and potentially losing some protein identifications) Validator can accommodate the data analysis.

We intend to provide Validator versions 1–3 both as a downloadable, open source program and as a Web-based tool for parsing and analyzing searched Mascot data. In addition, this approach is readily applied to other labeling schemes used for quantitative analysis, such as stable isotope labeling by amino acids in cell culture (SILAC) or ICAT. Thus, we intend to adapt the software to accommodate other stable isotope tags. Analysis will also be extended to other search platforms such as SEQUEST or X! Tandem.

This study raises the possibility of implementing a new approach to proteomics data acquisition and analysis to speed up and enhance protein identification based on iden-

tifying peptides "on the fly" during the LC-MS/MS run. Our data suggest that peptides might be readily identified, even in a complex sample, based on detecting pairs of precursor ions with a characteristic mass difference. Then MS/MS could be performed on both the heavy and light forms followed by comparison to detect shifting and non-shifting fragment ions. The lists of precursor ion masses and *b*- and *y*-ions determined from such a match could be used to generate sequence tags as done by Mann and Wilm (20) to directly identify each peptide and thus the protein. With such a strategy, protein identification in real time during the LC-MS/MS run is entirely feasible from a computational perspective. Toward these ends, we anticipate pursuing rapid recognition of $^{16}O/^{18}O$ pairs in raw LC-MS/MS data and interrogating pairs of fragmentation patterns to search for matching shifting and non-shifting ions.

In its current incarnation, our Validator software offers a simple and powerful tool to filter searched tandem mass spectrometry proteomics data. By applying the techniques outlined above, a list of high confidence peptide and protein identifications can be obtained within minutes, thus reducing the complexity of downstream proteomics analyses.

* This work was supported, in whole or in part, by National Institutes of Health Grants R01 GM60443 and R01 HG003864 (to S. K.). This work was also supported by a Department of Defense Breast Cancer Research Program multidisciplinary postdoctoral award (to K. K.) and a grant from the Cancer Research Foundation (to S. V.).

[S] The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

¶ To whom correspondence should be addressed: Computation Institute, The University of Chicago, 5640 S. Ellis RI405, Chicago, IL 60637. Tel.: 773-702-4303; Fax: 773-834-1329; E-mail: slv@uchicago.edu.

REFERENCES

1. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics. *Nature* **422,** 198–207
2. Hunt, D. F., Yates, J. R., 3rd, Shabanowitz, J., Winston, S., and Hauer, C. R. (1986) Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A.* **83,** 6233–6237
3. Lin, D., Tabb, D. L., and Yates, J. R., 3rd (2003) Large-scale protein identification using mass spectrometry. *Biochim. Biophys. Acta* **1646,** 1–10
4. Liu, T., Belov, M. E., Jaitly, N., Qian, W. J., and Smith, R. D. (2007) Accurate mass measurements in proteomics. *Chem. Rev.* **107,** 3621–3653
5. Nesvizhskii, A. I., Vitek, O., and Aebersold, R. (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat. Methods* **4,** 787–797
6. Taylor, J. A., and Johnson, R. S. (1997) Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **11,** 1067–1075
7. Frank, A., and Pevzner, P. (2005) PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* **77,** 964–973
8. Fischer, B., Roth, V., Roos, F., Grossmann, J., Baginsky, S., Widmayer, P., Gruissem, W., and Buhmann, J. M. (2005) NovoHMM: a hidden Markov model for de novo peptide sequencing. *Anal. Chem.* **77,** 7265–7273
9. DiMaggio, P. A., Jr., and Floudas, C. A. (2007) De novo peptide identification via tandem mass spectrometry and integer linear optimization. *Anal. Chem.* **79,** 1433–1446
10. Frank, A. M., Savitski, M. M., Nielsen, M. L., Zubarev, R. A., and Pevzner, P. A. (2007) De novo peptide sequencing and identification with precision mass spectrometry. *J. Proteome Res.* **6,** 114–123
11. Gu, S., and Chen, X. (2005) Precise proteomic identification using mass spectrometry coupled with stable isotope labeling. *Analyst* **130,** 1225–1231
12. Shevchenko, A., Chernushevich, I., Ens, W., Standing, K. G., Thomson, B., Wilm, M., and Mann, M. (1997) Rapid 'de novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **11,** 1015–1024
13. Tanner, S., Shu, H., Frank, A., Wang, L. C., Zandi, E., Mumby, M., Pevzner, P. A., and Bafna, V. (2005) InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal. Chem.* **77,** 4626–4639
14. Eng, J. K., McCormack, A. L., and Yates, J. R., 3rd (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5,** 976–989
15. Yates, J. R., 3rd, Eng, J. K., McCormack, A. L., and Schieltz, D. (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. *Anal. Chem.* **67,** 1426–1436
16. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20,** 3551–3567
17. Craig, R., and Beavis, R. C. (2004) TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20,** 1466–1467
18. Craig, R., and Beavis, R. C. (2003) A method for reducing the time required to match protein sequences with tandem mass spectra. *Rapid Commun. Mass Spectrom.* **17,** 2310–2316
19. Deutsch, E. W., Lam, H., and Aebersold, R. (2008) Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. *Physiol. Genomics* **33,** 18–25
20. Mann, M., and Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66,** 4390–4399
21. Tabb, D. L., Ma, Z. Q., Martin, D. B., Ham, A. J., and Chambers, M. C. (2008) DirecTag: accurate sequence tags from peptide MS/MS through statistical scoring. *J. Proteome Res.* **7,** 3838–3846
22. Tabb, D. L., Saraf, A., and Yates, J. R., 3rd (2003) GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. *Anal. Chem.* **75,** 6415–6421
23. Sunyaev, S., Liska, A. J., Golod, A., Shevchenko, A., and Shevchenko, A. (2003) MultiTag: multiple error-tolerant sequence tag search for the sequence-similarity identification of proteins by mass spectrometry. *Anal. Chem.* **75,** 1307–1315
24. Tabb, D. L. (2008) What's driving false discovery rates? *J. Proteome Res.* **7,** 45–46
25. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **74,** 5383–5392
26. Savitski, M., Nielsen, M. L., and Zubarev, R. A. (2005) New data base-independent, sequence tag-based scoring of peptide MS/MS data validates Mowse scores, recovers below threshold data, singles out modified peptides, and assesses the quality of MS/MS techniques. *Mol. Cell. Proteomics* **4,** 1180–1188
27. Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Posterior error probabilities and false discovery rates: two sides of the same coin. *J. Proteome Res.* **7,** 40–44
28. Elias, J. E., and Gygi, S. P. (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **4,** 207–214
29. Käll, L., Storey, J. D., MacCoss, M. J., and Noble, W. S. (2008) Assigning significance to peptides identified by tandem mass spectrometry using decoy databases. *J. Proteome Res.* **7,** 29–34
30. Choi, H., Ghosh, D., and Nesvizhskii, A. I. (2008) Statistical validation of peptide identifications in large-scale proteomics using the target-decoy database search strategy and flexible mixture modeling. *J. Proteome Res.* **7,** 286–292
31. Choi, H., and Nesvizhskii, A. I. (2008) Semisupervised model-based validation of peptide identifications in mass spectrometry-based proteomics. *J. Proteome Res.* **7,** 254–265
32. Ong, S. E., and Mann, M. (2005) Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1,** 252–262
33. Schnölzer, M., Jedrzejewski, P., and Lehmann, W. D. (1996) Protease-

catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis* **17,** 945–953
34. Scoble, H. A., and Martin, S. A. (1990) Characterization of recombinant proteins. *Methods Enzymol.* **193,** 519–536
35. Takao, T., Hori, H., Okamoto, K., Harada, A., Kamachi, M., and Shimonishi, Y. (1991) Facile assignment of sequence ions of a peptide labelled with 18O at the carboxyl terminus. *Rapid Commun. Mass Spectrom.* **5,** 312–315
36. Gray, W. R., and Del Valle, U. E. (1970) Application of mass spectrometry to protein chemistry. I. Method for amino-terminal sequence analysis of proteins. *Biochemistry* **9,** 2134–2137
37. Gray, W. R., Wojcik, L. H., and Futrell, J. H. (1970) Application of mass spectrometry to protein chemistry. II. Chemical ionization studies on acetylated permethylated peptides. *Biochem. Biophys. Res. Commun.* **41,** 1111–1119
38. Rose, K., Simona, M. G., Offord, R. E., Prior, C. P., Otto, B., and Thatcher, D. R. (1983) A new mass-spectrometric C-terminal sequencing technique finds a similarity between gamma-interferon and alpha 2-interferon and identifies a proteolytically clipped gamma-interferon that retains full antiviral activity. *Biochem. J.* **215,** 273–277
39. Cagney, G., and Emili, A. (2002) De novo peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging. *Nat. Biotechnol.* **20,** 163–170
40. Goodlett, D. (2003) Stable isotopic labeling and mass spectrometry as a means to determine differences in protein expression. *Trends Analyt. Chem.* **22,** 282–290
41. Goodlett, D. R., Keller, A., Watts, J. D., Newitt, R., Yi, E. C., Purvine, S., Eng, J. K., von Haller, P., Aebersold, R., and Kolker, E. (2001) Differential stable isotope labeling of peptides for quantitation and de novo sequence derivation. *Rapid Commun. Mass Spectrom.* **15,** 1214–1221
42. Gu, S., Pan, S., Bradbury, E. M., and Chen, X. (2003) Precise peptide sequencing and protein quantification in the human proteome through in vivo lysine-specific mass tagging. *J. Am. Soc. Mass Spectrom.* **14,** 1–7
43. Lee, Y. H., Han, H., Chang, S. B., and Lee, S. W. (2004) Isotope-coded N-terminal sulfonation of peptides allows quantitative proteomic analysis with increased de novo peptide sequencing capability. *Rapid Commun. Mass Spectrom.* **18,** 3019–3027
44. Qin, J., Herring, C. J., and Zhang, X. (1998) De novo peptide sequencing in an ion trap mass spectrometer with 18O labeling. *Rapid Commun. Mass Spectrom.* **12,** 209–216
45. Zhang, N., Aebersold, R., and Schwikowski, B. (2002) ProbID: a probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. *Proteomics* **2,** 1406–1412
46. Pratt, J. M., Robertson, D. H., Gaskell, S. J., Riba-Garcia, I., Hubbard, S. J., Sidhu, K., Oliver, S. G., Butler, P., Hayes, A., Petty, J., and Beynon, R. J. (2002) Stable isotope labelling in vivo as an aid to protein identification in peptide mass fingerprinting. *Proteomics* **2,** 157–163
47. Potthast, F., Gerrits, B., Häkkinen, J., Rutishauser, D., Ahrens, C. H., Roschitzki, B., Baerenfaller, K., Munton, R. P., Walther, P., Gehrig, P., Seif, P., Seeberger, P. H., and Schlapbach, R. (2007) The Mass Distance Fingerprint: a statistical framework for de novo detection of predominant modifications using high-accuracy mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **854,** 173–182
48. Bandeira, N., Tsur, D., Frank, A., and Pevzner, P. A. (2007) Protein identification by spectral networks analysis. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 6140–6145
49. Savitski, M. M., Nielsen, M. L., and Zubarev, R. A. (2006) ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol. Cell. Proteomics* **5,** 935–948
50. Heller, M., Mattou, H., Menzel, C., and Yao, X. (2003) Trypsin catalyzed 16O-to-18O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers. *J. Am. Soc. Mass Spectrom.* **14,** 704–718
51. Kolker, E., Hogan, J. M., Higdon, R., Kolker, N., Landorf, E., Yakunin, A. F., Collart, F. R., and van Belle, G. (2007) Development of BIATECH-54 standard mixtures for assessment of protein identification and relative expression. *Proteomics* **7,** 3693–3698
52. Elias, J. E., Haas, W., Faherty, B. K., and Gygi, S. P. (2005) Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat. Methods* **2,** 667–675
53. Mirgorodskaya, O. A., Kozmin, Y. P., Titov, M. I., Körner, R., Sönksen, C. P., and Roepstorff, P. (2000) Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (18)O-labeled internal standards. *Rapid Commun. Mass Spectrom.* **14,** 1226–1232
54. Peng, J., Elias, J. E., Thoreen, C. C., Licklider, L. J., and Gygi, S. P. (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2,** 43–50
55. Qian, W. J., Monroe, M. E., Liu, T., Jacobs, J. M., Anderson, G. A., Shen, Y., Moore, R. J., Anderson, D. J., Zhang, R., Calvano, S. E., Lowry, S. F., Xiao, W., Moldawer, L. L., Davis, R. W., Tompkins, R. G., Camp, D. G., 2nd, and Smith, R. D. (2005) Quantitative proteome analysis of human plasma following in vivo lipopolysaccharide administration using 16O/18O labeling and the accurate mass and time tag approach. *Mol. Cell. Proteomics* **4,** 700–709
56. Sakai, J., Kojima, S., Yanagi, K., and Kanaoka, M. (2005) 18O-labeling quantitative proteomics using an ion trap mass spectrometer. *Proteomics* **5,** 16–23
57. Yao, X., Freas, A., Ramirez, J., Demirev, P. A., and Fenselau, C. (2001) Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.* **73,** 2836–2842

Stable-Isotope Labeling for Protein Quantitation by Mass Spectrometry

Kolbrun Kristjansdottir and Stephen J. Kron*

Department of Molecular Genetics and Cell Biology and Ludwig Center for Metastasis Research, The University of Chicago, Chicago, IL 60637, USA.

ABSTRACT: Mass spectrometry has become a routine instrument to identify proteins and peptides from simple or complex samples. Although identification can be confidently determined from a single experiment, quantitation requires multiple replicates and careful analysis. Alternatively, stable isotopes can be used to obtain relative quantitation of proteins and peptides from fewer replicates. Conventionally, half of a sample is labeled with stable isotope and mixed with the other half of unlabeled sample. The mixed sample is analyzed by mass spectrometry and because the stable isotope does not change the chemical properties of the peptide, the intensities of the unlabeled and labeled peptide can be directly compared. Absolute quantitation is obtained by adding a known amount of stable isotope labeled peptide or protein and comparing to an unlabeled counterpart. Stable isotope labeling methodologies can be divided into three categories: Chemical, enzymatic and metabolic. Here we provide an up-to-date review comparing the benefits and drawbacks of all three stable isotope labeling methodologies and briefly describe quantitation software solutions. In addition to quantitation, stable isotopes have also been used to identify post-translational modifications in proteins, identify components of DNA-protein and protein-protein complexes and to distinguish background contaminants from experimental results. Finally, we describe how fragmentation patterns from stable isotope labeled peptide and unlabeled peptides can improve peptide and protein identification and validation.

Keywords: Mass spectrometry, quantitation, stable isotope, isobaric, labeling, chemical, metabolic, enzymatic, iTRAQ, SILAC, ICAT, proteomics, software.

INTRODUCTION

The combination of complete proteolysis, peptide separation by reverse-phase liquid chromatography, and detection by electrospray ionization and tandem mass spectrometry (LC-MS/MS, reviewed in [1] offers a powerful approach to comprehensive detection of the proteins and their modifications in complex samples. It has long been recognized that along with making the mass measurements of peptide ions and their fragments required for identification, LC-MS/MS instruments also record peptide ion intensities, offering the potential for direct measurement of peptide concentration and thereby protein abundance. However, the extent of ionization of peptides by electrospray ionization is dependent on peptide sequence and modification, elution conditions, complexity of the sample and other factors. As a result, the absolute intensities of ions derived from non-identical peptides cannot provide accurate or direct quantitation. Approaches such as peptide ion chromatogram extraction and spectral counting have been developed to obtain relative quantitation of protein abundance [2-10]. These approaches, collectively termed "label-free" quantitation, require extensive analysis of reference samples and/or significant data redundancy, often requiring many hours of mass spectrometry time per sample. Although highly promising, label-free approaches remain impractical for users lacking access to dedicated mass spectrometry instrumentation and advanced informatic approaches.

Stable-isotope labeling provides an attractive alternative to label-free approaches. A stable-isotope labeled peptide and its unlabeled counterpart have the same chemical formula and structure and thus (almost) identical chemical properties, such that they are expected to elute together from reverse phase and then ionize and fragment identically in the mass spectrometer, yet can be followed independently based on their mass differential. Combining the light (unlabeled) and heavy-isotope labeled peptides in one sample allows for direct comparison of ion intensities. In principle, this offers highly accurate relative quantitation and avoids the need for significant data redundancy. Background peaks are readily distinguished from "real" peptides insofar as the "real" peptides are represented by both light and heavy forms with a characteristic mass offset. With these and other advantages, stable-isotope labeling would appear to satisfy the criteria for an ideal quantitative mass spectrometry strategy. However, challenges remain to be addressed before stable-isotope quantitation becomes a straightforward, robust and reliable approach accessible both to non-experts and users of service laboratories. Stable-isotope labeling of peptides/proteins can be performed using chemical, enzymatic or metabolic methods and each one of these methods has been reviewed individually [11-13]. Here, we provide an up-to-date and critical analysis comparing the benefits and drawbacks of all three stable-isotope labeling methodologies and explore the state-of-the-art, caveats and concerns and emerging new applications of these powerful approaches.

*Address correspondence to this author at the Ludwig Center for Metastasis Research, 924 E. 57th St., Chicago, IL 60637, USA; Tel: 1-773-834-0250; Fax: 1-773-702-4394; E-mail: skron@uchicago.edu

1570-1646/10 $55.00+.00

Several articles have highlighted the importance of high mass accuracy for both protein identification and quantitation [14-16]. Readers are likely to see increasing access to commercially available reliable, high mass accuracy, high-resolution mass spectrometers capable of high-throughput tandem mass spectrometry, that obviate many challenges for quantitative analysis of data from low-resolution mass spectrometers. Thus, this review will only cover the use of stable isotopes in high-resolution mass spectrometers.

STABLE ISOTOPES IN NATURE: ISOTOPIC ENVELOPE FOR PEPTIDES

Isotopes of an element share the same number of protons but distinct atomic weights due to different numbers of neutrons. Only a small number of possible isotopes, limited to those with the right numbers of neutrons to balance electrostatic and strong binding forces in the atomic nucleus, are sufficiently stable to be non-radioactive and thereby accumulate in Nature. Though most possible isotopes formed *via* fusion or fission reactions are profoundly unstable and decay instantaneously, some remain intact with half-lives from a few minutes to millenia and are considered radioactive. Labeling of proteins with atomic isotopes to follow changes in protein abundance or modification *in vivo* has a long history. Conventional methods dependent on radioactive isotopes used as tracers remain useful to detect rates of protein synthesis and degradation, typically *via* pulse-chase approaches. Here, a short period of metabolic incorporation of labeled amino acids synthesized with a high specific activity of ^{3}H, ^{14}C or ^{35}S isotopes leads to transient labeling of proteins. The kinetics of translation, maturation and/or proteolysis can be followed using methods such as immunoprecipitation, gel electrophoresis and detection of beta particle emission by autoradiography. Given the low natural background, the beta decay events can be readily detected with high signal-to-noise. Thus, even trace radioactive isotope labeling is sufficient for sensitive detection and precise quantitation.

However, for stable-isotope labeling, trace incorporation is not sufficient. Nearly all of the elements that are common in proteins including carbon, hydrogen, oxygen, nitrogen, and sulfur, have two or more isotopes with measurable abundance in Nature, with the lightest of these present in greater abundance than the others. For example, carbon is found in three forms in nature, the predominant stable "light" isotope ^{12}C (98.89%), a stable heavy isotope of ^{13}C (1.11%) and a radioactive heavy isotope of ^{14}C (trace amounts). Nitrogen is found in two forms: light ^{14}N (99.63%) and a stable heavy isotope of ^{15}N (0.37%). Oxygen is present predominantly as ^{16}O (99.76%), but ^{17}O (0.04%) and ^{18}O (0.20%) are comparatively common stable isotopes. Sulfur is present as ^{32}S (94.93%), ^{33}S (0.76%), ^{34}S (4.29%) and ^{36}S (0.02%). Finally, hydrogen is predominantly ^{1}H (99.98%), but ^{2}H (deuterium, 0.02%) and traces of ^{3}H (tritium) are present. In general, heavy isotopes display a kinetic isotope effect on chemical reactions, slowing reaction rates and leading to a comparative underrepresentation in complex molecules, as are made by living organisms. However, save for deuterium, when incorporated into amino acids, the different isotopes are (mostly) indistinguishable to biological organisms and are incorporated non-discriminately into proteins. Since carbon and nitrogen are the most common atoms in peptides and ^{13}C and ^{15}N are abundant in nature, they, along with ^{34}S, are the predominant heavy isotopes naturally present in proteins. As a result, instead of each tryptic peptide having a single mass, mass spectrometry spectra reveals a collection of different masses in proportions that reflect the natural abundance of isotopes. Fig. (**1**) shows a collection of peaks all representing isotopic forms of a single peptide, termed an isotopic envelope. A pattern of four major peaks with a characteristic pattern of intensity are detected at 790.89, 791.39, 791.89, 792.39, 792.89 (mass/charge). The first peak at 790.89 m/z is designated as the monoisotopic ion of the 2^{+} charged peptide, representing the form that corresponds to the chemical formula and contains only the common isotopes ^{1}H, ^{12}C, ^{14}N, ^{16}O, etc. The second peak at 791.39 is 0.5 m/z units higher than the monoisotopic peak. This corresponds to an ~1 Dalton increase in mass due to the presence of a single stable isotope. Most of this peak is due to peptide isotopologues carrying a single ^{13}C, and the m/z shift corresponds to the mass of the additional neutron divided by the charge of the peptide, 2. The third peak at 791.89 represents the peptide with two stable isotopes, often a pair of ^{13}C's or one ^{13}C and one ^{15}N or a single ^{34}S or ^{18}O, divided by the charge of the peptide, 2, to give an m/z shift of 1 and so on. Note that each of these peaks includes forms with slightly different masses, due to the individual mass defects (binding energy) of the different stable-isotope nuclei. The intensity of each peak is defined by a combination of the abundance of specific isotopes in Nature and the occurrence of each element in the peptide. For peptides of roughly twenty residues or greater, the +1 peak will have a greater abundance than the monoisotopic form and for most proteins (e.g. > 100 residues), the monoisotopic form cannot be detected. Taken together, the pattern and intensity of each isotopic envelope of peptide sequences can be predicted (described in [17]). Several open source software tools are available to predict isotopic envelopes such as Isotopica [18] and Envelope [19]. Isotopic envelopes can be sufficiently resolved on high-resolution and high precision mass spectrometers to provide an additional criterion that can enhance the confidence of peptide identifications.

In general, the challenges of stable isotope-based quantitation were addressed long ago in the development of analysis methods for isotope-dilution mass spectrometry as reviewed in [20]. However, the classical approaches do not scale directly to analysis of macromolecules. As a consequence of the high natural abundance of stable isotopes and the large number of atoms in each peptide, it follows that in order to obtain sufficient signal-to-noise to distinguish labeled from unlabeled forms of a peptide, the stoichiometry of artificial incorporation must be relatively high in the heavy "reference" sample. In practice, a shift of 2 Da or more between the most abundant isotopologues of the "light" and "heavy" labeled peptides is required to obtain satisfactory quantitation.

LABELING OF PEPTIDES AND/OR PROTEINS USING STABLE ISOTOPES

In the past decade, several effective methods for stable-isotope labeling of peptides and proteins have been reported and used to determine the relative abundance of proteins using mass spectrometry [21-23]. Common to all these tech-

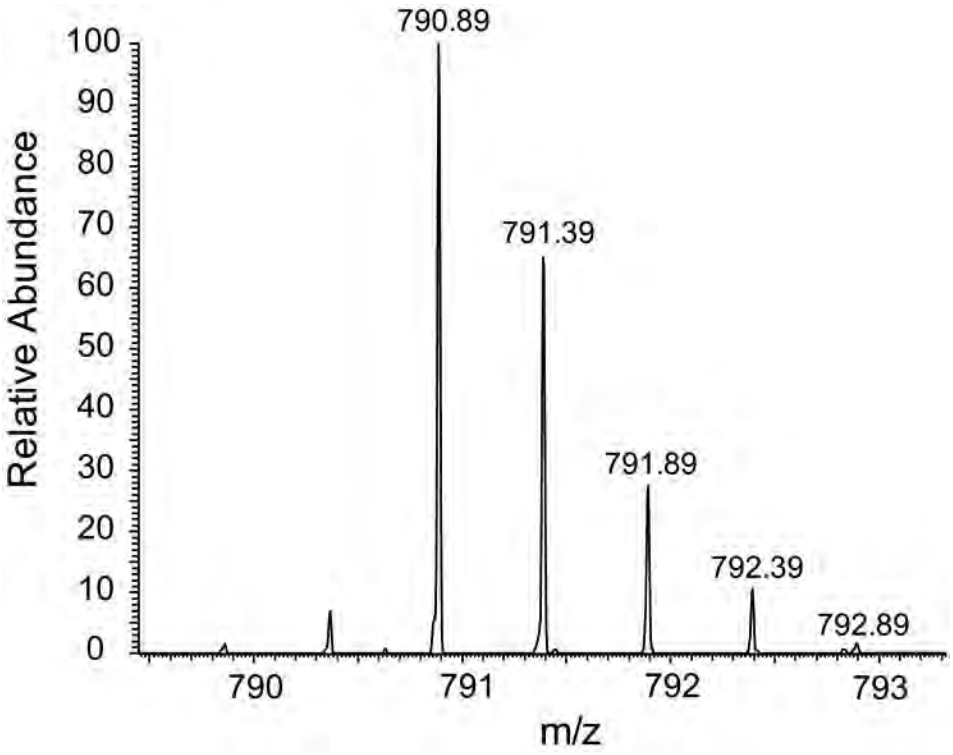

Fig. (1). MS1 spectrum showing an isotopic envelope of a 2+ charged peptide. Four forms are detected. 790.89 m/z represents the peptide with no minor stable-isotopes. Peaks at 791.39, 791.89, 792.39, 792.89 correspond to the peptide containing one, two, three and four minor stable-isotopes, respectively.

niques is metabolic, enzymatic or chemical incorporation of a labeling moiety being enriched with heavy stable isotopes such as deuterium, ^{13}C, ^{18}O, or ^{15}N. The isotope-labeled sample is mixed with an equal amount of unlabeled sample to provide relative quantitation (heavy/light ratio). An overview of stable-isotope labeling methods is presented in Table **1** and Fig. (**2**).

METABOLIC LABELING

Uniform metabolic labeling of organisms with heavy isotopes dates from shortly after the discovery of heavy water in the early 1930's and is found in a number of applications, including increasing the sensitivity and resolution of NMR. Indeed, stable-isotope labeled nutrients derived from micro-organisms cultured in ^{2}H, ^{13}C and/or ^{15}N have long been commercially available and comparatively inexpensive. Metabolic labeling for quantitation was first introduced to proteomics by the Chait group [24] who grew yeast on a commercial rich media derived from ^{15}N-enriched algal hydrolysate and measured relative abundance of phosphopeptides in the light and heavy samples by MALDI mass spectrometry. Analogous approaches have been applied with a number of organisms including worms and flies, culminating with the work of Wu *et al*. [25] who metabolically labeled a rat by feeding with ^{15}N-enriched algae to produce tissue-specific internal standards for global quantitative proteomic analysis. A disadvantage of this approach is that the distribution of isotopic forms for each peptide depends on the amino acid composition, complicating quantitative analysis and manual validation.

Stable-Isotope Labeling in Cell Culture (SILAC)

Currently, the most widely used metabolic labeling approach for protein quantitation is SILAC, stable-isotope labeling with amino acids in cell culture [26-29]. When cells are grown for several doublings in tissue culture with a stable-isotope labeled form of an essential amino acid (e.g. lysine) as the sole source and at a small excess, it is incorporated into newly synthesized proteins until all proteins are homogeneously labeled (Fig. **2**, right panel). Although any of the 20 naturally occurring amino acids could be used as a precursor for labeling, several factors argue for specific amino acids being selected for SILAC (reviewed in [30]). The most common is leucine, followed by lysine, arginine, and to a lesser extent serine, glycine, histidine, methionine, valine, and tyrosine. The most common isotopes in SILAC are ^{13}C and ^{15}N, since they demonstrate less kinetic isotope effect than ^{2}H and do not change the elution profiles of labeled peptides in reverse phase HPLC chromatography [31-33].

Trypsin is the most common protease used in proteomics, cleaving carboxyl-terminal to lysine and arginine residues. Therefore, each tryptic peptide is predicted to contain either a single, carboxyl terminal lysine or arginine. Growing cells in the presence of stable-isotope labeled arginine and lysine as the sole source, followed by trypsin digestion, yields tryptic peptides terminated by a stable-isotope labeled amino acid. With a mass difference of typically 4 to 10 due to labeling of the single terminal lysine or arginine, most pairs of peptides can be easily recognized by their offset envelopes of isotopic species (Fig. **3**).

Table 1. An Overview of Methods for Stable-Isotope Labeling of Peptides and/or Proteins

	SILAC	Isobaric Tags	ICAT	^{18}O Labeling
Type of labeling	Metabolic	Chemical	Chemical	Enzymatic
Time of labeling	First step (cell growth)	Middle step (peptide labeling)	Middle step (protein labeling)	Final step (peptide labeling)
Sample type	Sample that can grow in cell culture (Cell lines, yeast, bacteria)	Any	Any	Any
Post-label fractionation	Peptide and protein separation	Peptide separation	Peptide and protein separation	Only peptide separation
Labeling target	Proteins, selected amino acid	N-terminal of peptides and lysine side chain	Peptides containing cysteines	C-terminus of all peptides
Sample number	Usually 2 (Up to 5)	4 or 8	2	2
MS level	MS1	MS2 (MS/MS)	MS1	MS1
Sample complexity	Increased	Same	Increased	Increased

The advantages of SILAC using lysine and arginine as the labeled amino acids include the ease of complete (100%) labeling and complete coverage of each protein save for its C-terminal peptide. That trypsin, even after "complete" digestion, predictably fails to cleave at some lysine and arginine residues (e.g. post-translationally modifed lysine or arginine, specific sequence contexts) somewhat complicates analysis, but does not prevent quantitation. Stable-isotope labeled amino acids (e.g. Cambridge Isotopes) and several types of SILAC tissue culture media including DMEM, RPMI and IMEM (Thermo Scientific Pierce, Invitrogen) are commercially available. SILAC is limited to organisms that can be grown on defined media. This is straightforward for cell lines, bacterial and yeast cells, but precludes most animal or human studies. Finally, SILAC is most straightforward when experiments consist of 2 samples, a control (heavy) and treatment (light). However, recent studies have combined samples each labeled with different isotopic forms of the same amino acid, i.e. Arg, $^{13}C_6$ Arg, $^{13}C_6$-$^{15}N_4$ Arg, etc., to obtain comparative quantitation of three [34] to five conditions [35].

STABLE-ISOTOPE LABELING USING CHEMICAL METHODS

Incorporation of stable isotopes into peptides or proteins *via* chemical reaction offers flexibility in sample types, including tissues and bodily fluids. Common strategies include

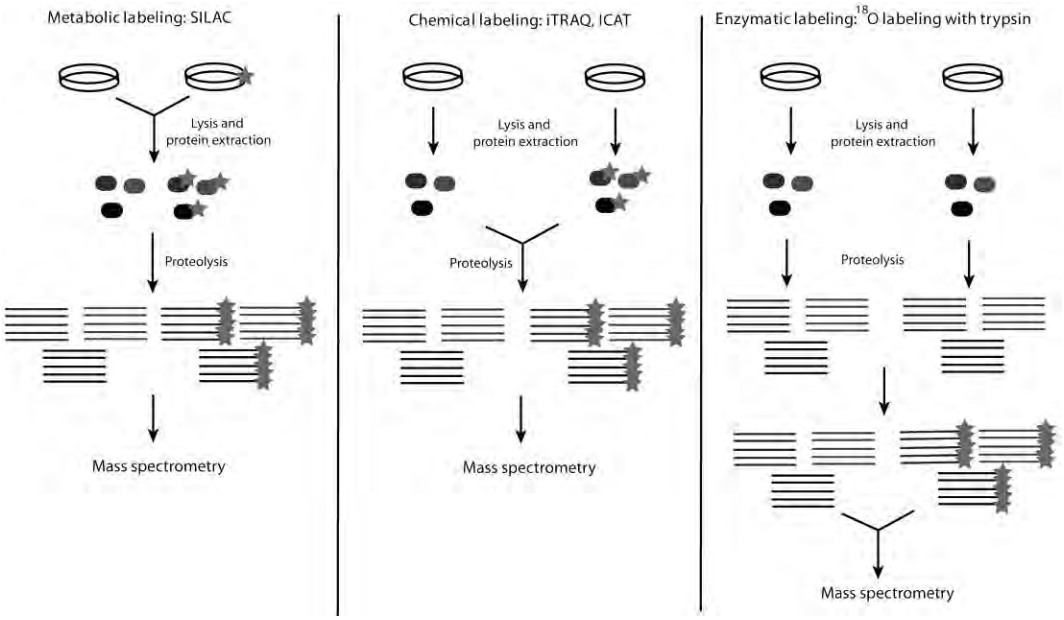

Fig. (2). An overview of stable-isotope labeling methods. The three types of labeling are: chemical (iTRAQ, ICAT), enzymatic (^{18}O labeling with proteases, f.ex. trypsin) and metabolic (SILAC). Stars indicate presence of stable-isotope. Labeling can occur at the cellular level (metabolic labeling), at the protein level (chemical labeling) or at the peptide level (enzymatic and chemical labeling). Stable-isotope labeled peptides are identified and quantified in the final step, mass spectrometry analysis.

targeting the N- or C-terminal or any of the chemically reactive amino acid side chains of peptides or proteins (Fig. **3**, left panel). Chemical methods are not restricted in the size of the stable-isotope reagent and can be synthesized to include cleavable modules and/or affinity tags for isolation of a targeted subset of the proteome. Recent methodologies include labeling with large isobaric (identical mass but distinct chemistry and/or isotopic distribution) tags that are cleaved during peptide fragmentation releasing marker ions. Comparing the intensity of these marker ions at the MS/MS level provides relative quantitation. Disadvantages of chemical approaches include sample-to-sample inconsistency due to incomplete labeling on target sites and competing side reactions that can modify secondary sites.

Examples of Chemical Labeling Methods

a) Isotope-Coded Affinity tag (ICAT)

One of the first commercialized stable-isotope tagging reagents is Isotope-coded affinity tag (ICAT, Applied Biosystems) [21, 36-39]. In ICAT, a pair of light and heavy reagents target cysteines on peptides, adding a linker and a biotin tag for affinity purification. The linker region of the heavy reagent contains stable isotopes whereas the light reagent contains no stable isotopes. Proteins from the samples to be examined are denatured and labeled with heavy or light reagents and then mixed and proteolyzed. The biotinylated peptides are purified using avidin affinity reagents, allowing for stringent washing that lowers background binding.

The main advantage of this method is that it enriches peptides containing the relatively rare amino acid cysteine, thereby significantly reducing the complexity of the peptide mixture and increasing the dynamic range of mass spectrometry analysis. The downside is that only peptides and proteins containing cysteines are identified, giving low overall coverage. As a result quantitation becomes less accurate since few peptides are obtained from each protein. Finally, ICAT is limited to comparing two samples.

The ICAT approach has been widely used since its introduction in 1999 [21, 36-39]. ICAT reagents have been commercialized and are available from Applied Biosystems. Several global quantitation experiments have been performed using the ICAT approach including the original paper where protein expression in yeast *Saccharomyces cerevisiae* was compared using either ethanol or galactose as a carbon source [21]. Other ICAT studies include identification of proteins regulated by the Myc oncoprotein [38] by comparing the protein expression patterns between myc-null and myc expressing cells and identification of proteins regulated by interferon treatment in human liver cells [39].

b) Other Cysteine Labeling Methods

Several other methods have been developed for chemical labeling of cysteines including HysTag [40] and acrylamide labeling [41]. HysTag is a 10-mer derivatized peptide, which consists of an affinity ligand (His6-tag), a tryptic cleavage site, a Ala-9 residue that contains either four (D_4) or no (D_0) deuterium atoms, and a thiol-reactive group targeting cysteines. The HysTag peptide is preserved in Lys-C digestion of proteins and allows subsequent charge-based selection of cysteine-containing peptides. To remove the HysTag, subsequent tryptic digestion reduces the labeling group to a dipeptide, which does not hinder effective MS/MS fragmentation [40]. HysTag has many of the same advantaged and disadvantages of ICAT.

The second method involves alkylation of cysteines of intact proteins with acrylamide [41]. While cysteine alkylation with acrylamide *via* Michael addition is an undesired reaction that frequently occurs during polyacrylamide gel electrophoresis [42], several features make it a useful tagging approach for quantitative analysis with stable isotopes. First, because of its small size and hydrophilic nature, the acrylamide moiety does not introduce significant mass shift or charge changes in the protein and does not negatively affect protein solubility. Second, cysteine labeling is facile allowing for complete labeling. Finally, the reagents are relatively inexpensive, making it practical to perform experiments starting with large amounts of protein as needed for extensive fractionation and in-depth analysis [41]. The disadvantages of acrylamide labeling are, as with other cysteine labeling reagents, that only cysteines are labeled and only peptides containing cysteines can be quantified. However, as opposed to ICAT, the acrylamide method does not include a cysteine peptide enrichment step. Finally, the mass shift is small, 3 Dalton, resulting in some overlap between the isotope envelopes of light and heavy peptides.

c) Isobaric Tags

The chemical labeling technique iTRAQ (Isobaric tags for relative and absolute quantitation) developed by Pappin [43] allows for quantitative comparison of up to 8 conditions without increasing sample "complexity". This method differs from the previous methods in that the quantitation is performed at the MS/MS level. The iTRAQ reagent consists of a reporter group, a balance group and a reactive group that reacts with lysine side chains and N-terminal groups of peptides. In the original 4-component version, the reporter group masses are 114, 115, 116 or 117 Da and the balance group masses are 31, 30, 29 or 28 Da to ensure that the combined mass remains constant at 145 Da. Briefly, a control and three treated samples are labeled individually with one of the iTRAQ reagents. The samples are then combined. Given that each isobaric tag has the same minor effect on the elution properties of the peptide, the four labeled versions of each peptide are indistinguishable in MS1 and are selected to fragment within a single MS/MS scan. During collision-induced fragmentation (CID), the reporter group ions (114, 115, 116 and 117 Da) break away from the backbone peptides, without preventing the fragmentation at peptide bonds needed for peptide identification. Relative quantitation for each of the treatment conditions being studied is obtained by comparing the intensities of the reporter group fragments. Isobaric tags have been commercialized. 4- and 8-component iTRAQ kits (reporter groups of 113, 114, 115, 116, 117, 118, 119 and 121 Da) are available from Applied Biosystems. Tandem Isobaric Mass Tag (TMT) kits with two or six components that work by a similar principle are available from Thermo Scientific.

The primary benefit of isobaric tags over ICAT and related approaches is that labeling does not increase the complexity of the mixture at the MS1 level, potentially resulting

in higher proteome coverage. Among downsides to isobaric tagging are that it is limited to instruments that can provide good MS/MS spectra in the 100-120 Da range, such as the QSTAR Quadrupole Time-of-Flight instrument (ABI). Recently, pulsed Q dissociation (PQD) has made it possible to detect the low mass isobaric tag reagent fragments on linear ion trap instruments including the LTQ-Orbitrap (Thermo) [44, 45].

As with other chemical labeling methods, complete labeling and removal of derivatization byproducts is required. Global quantitation experiments have been performed using the iTRAQ approach including time resolved monitoring of kinase reactions [46], comparison of organelle proteomes [47 and monitoring of protein expression changes as cancer cells acquire increasing metastatic potential [48]. Combining quantitation with phosphoproteomics, Aebersold *et al*. [49] recently described an iTRAQ method to simultaneously identify components and phosphorylation sites of protein complexes.

STABLE-ISOTOPE LABELING USING ENZYMATIC METHODS

Protease-Mediated ^{18}O Exchange

A third method of stable isotope labeling involves enzymatic transfer of ^{18}O from water to the carboxyl terminal of peptides by an oxygen exchange reaction [23, 50-53]. Several enzymes are capable of this reaction including bovine trypsin, Lys-C or Arg-C, with trypsin being the most commonly used. Trypsin digestion is the most common method of sample preparation before mass spectrometry and therefore, incubation of peptides with trypsin in ^{18}O enriched water is a straightforward addition to the workflow. Because the labeling occurs at the last step, the experimental and control sample must be kept separate during lysis, any protein enrichment and digestion.

Although ^{18}O labeling is possible during digestion, the separate labeling exchange reaction after proteolysis is preferable. Advantages include small volume labeling (decreasing the volume of $H_2{}^{18}O$ required), ready use of immobilized trypsin to reduce back-exchange and separate optimization of digestion and labeling [23].

Typically, tryptic ^{18}O labeling is performed after a complete digestion in ^{16}O water. One sample is then subjected to trypsin exchange in regular water (^{16}O sample) and the other in $H_2{}^{18}O$ water, resulting in the incorporation of two ^{18}O atoms to the C-terminus of the peptide (^{18}O sample) (Fig. **3**, middle panel). The samples are then mixed and the ^{16}O and ^{18}O forms of each peptide elute together from the HPLC as pairs of ions, which are identical save for their carboxyl ends. Similar to SILAC and ICAT, the relative abundance of peptides can be inferred based on the relative intensity between the "light" ^{16}O and "heavy" ^{18}O ions in the MS1 spectra.

The overall advantages of protease-mediated ^{18}O exchange are that essentially any sample can be labeled, labeling introduces no chemical changes to the peptides, and the work flow is simple and inexpensive. The disadvantages include that only 2 samples can be labeled and that samples must be kept separate throughout the lysis, enrichment and proteolysis steps, potentially introducing errors due to differences in sample handling. Another disadvantage is that labeling is not as reproducible as some chemical methods, as the exchange reaction is highly sequence specific, and relies heavily on the purity of the $H_2{}^{18}O$, the labeling time, buffer and temperature and the amount and activity of trypsin used.

COMMON CONCERNS ABOUT LABELING

A critical component to stable-isotope labeling, using chemical, enzymatic or metabolic methods, is achieving complete labeling. It is worth the effort to spend time optimizing and testing a labeled sample before starting an ex-

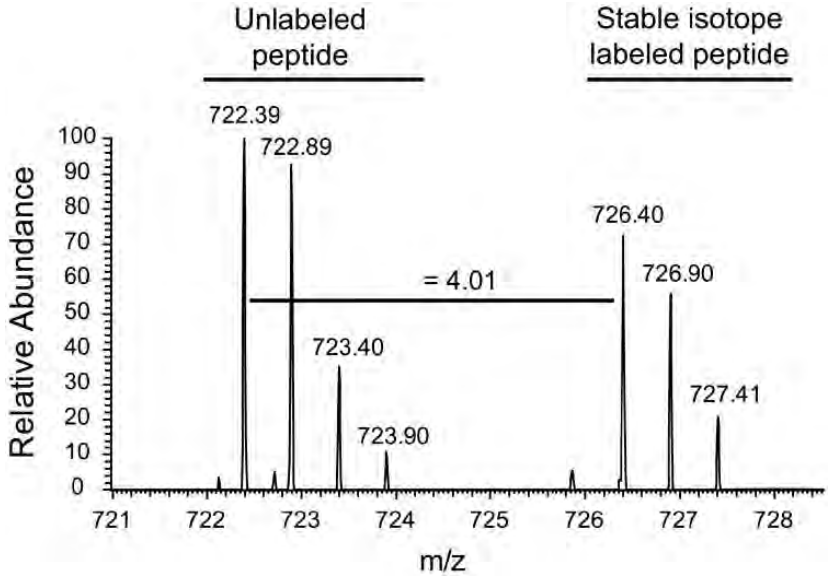

Fig. (3). MS1 spectra of unlabeled and stable-isotope labeled peptide. The sample was SILAC labeled with $^{13}C_6$,$^{15}N_2$-Lysine. The difference between the two peptides is 4.01 dalton corresponding to a single $^{13}C_6$,15N2-Lysine (8.014 dalton) corrected for peptide charge (2 charged peptide).

periment. Although calculations can be done to normalize samples to extent of labeling, downstream analysis will be greatly simplified if labeling is complete. Unfortunately, even with optimization to achieve stoichiometric labeling of the majority of peptides, each of the methods is subject to one or more artifacts, resulting in a subset of peptides that display partial or unexpected labeling, thereby confounding analysis.

All of the above mentioned methods of labeling, except for isobaric tags, result in generation of peptide pairs at the MS1 level, where the light and heavy peptides are separated by a predictable number of mass units. If the mass difference is small, the natural isotope distribution of the light form will overlap with the monoisotopic peak of heavy form, frustrating quantitiation. Trypsin-mediated ^{18}O exchange yields a 4 Da mass difference that leads to challenging quantitation of higher charged peptides and peptides over 20 residues, particularly if the labeling is incomplete. Indeed, incorporation of a single ^{18}O is common, leading to a mass difference of only 2 Da. In turn, even though acrylamide labeling is typically complete, it offers as little as a 3 Da mass shift. Although it is possible to deconvolute such overlapping distributions and quantify the heavy and light peaks, this is a complex and iterative process, requires high quality data, and is tedious. Thus, most commercial labeling reagents (SILAC, ICAT) are generated to have $\geq$ 4 Da mass difference and avoid this complication.

Finally, technical and biological replicates should be included to identify problems with labeling, quantitation and analysis. Potential problems include differences in sample handling, cell growth, labeling procedure and/or quantitation.

SOFTWARE FOR QUANTITATIVE ANALYSIS OF STABLE-ISOTOPE LABELED SAMPLES

Although manual analysis is possible, automated identification and quantitation of stable-isotope labeled peptides is far more practical but requires post-processing with specialized software. Software selection is based on the type of mass spectrometer used to generate the data, which varies by mass spectrometer and detector technology (time-of-flight, ion trap, Orbitrap, ICR, etc.) and manufacturer (Thermo, Agilent, Waters, Bruker, Applied Biosystems). Some of the software can handle data from several types of proprietary input files and others can only handle a single type. For this reason, we will not go into specifics of each software tool but rather list of some of the most popular software tools available. Currently available software are described in detail in recent reviews [54, 55].

Software for Quantitation at the MS1 Level

Mass spectrometry manufacturers often provide proprietary software solutions for quantitation. Examples include Bioworks (Thermo-Finnigan), Peakpicker (Applied Biosystems) and WARP-LC™ 1.1 (Bruker). Several open-source software tools are available including AYUMS developed by Miyano *et al.* [56], ProRata developed by Hettich *et al.* [57], and Mascot File Parsing and Quantification (MFPaQ) developed by Monsarrat *et al.* [58]. Compilations of software are available including Trans Proteomic Pipeline (TPP) developed at the Institute for Systems Biology (ISB) in Seattle. Modules for quantitation include XPRESS [59] and ASAPratio [60]. The ISB tools have been incorporated into Computational Proteomics Analysis System (CPAS), a suite of database and analysis tools, which manages proteomics-based experimental workflows and integrates database search algorithms [61]. CPAS was originally developed in the Fred Hutchinson Cancer Research Center but is now distributed as part of the Labkey Server, an open-source project managed by the Labkey Software Foundation. Most recently, an open-source integrated suite of algorithms specifically developed for quantitation of high-resolution MS data, termed MaxQuant, was developed by Matthias Mann's group [62]. Taking into account likely sources of error as described above, none of these software packages provides reliable quantitation without some manual validation.

Software for Quantitation at the MS/MS Level

Quantitation software for isobaric tags includes commercially available solutions such as ProteinPilot and ProQuant from Applied Biosystems, Spectrum Mill from Agilent, Proteome Discoverer from Thermo Scientific and Scaffold Q+ from Proteome Software. Open-source software includes Libra, a software module used within the Trans Proteomic Pipeline (TPP).

COMMON CONCERNS ABOUT QUANTITATION AND SUGGESTIONS TO IMPROVE QUALITY OF DATA

Despite the broad range of available software, manual validation is often necessary to confirm each peptide quantitation (and identification). Inaccurate or ambiguous results are almost certain where too few peptides can be quantified from a protein or where the standard deviation or p-value between multiple quantified peptides from a protein is not statistically significant. High-abundance proteins that yield ratios close to 1:1 have the highest confidence levels but provide little or no biological insight. As with any mass spectrometry experiment, low-abundance proteins are difficult to study because of the limited dynamic range. If peptides are close to the detection limit of the mass spectrometer, they can flicker in and out of the spectra making quantitation uncertain. Some of these difficulties cannot be addressed without fractionating and/or normalizing the sample, which are subject to their own costs and artifacts. We recommend obtaining both biological and technical replicates and/or reversing the labeling to obtain higher confidence in protein ratios.

Finally, if the sample is too complex (too many peptides are in the sample), overlapping peptide spectra can occur and bring about errors in peptide quantitation both in MS1 and MS/MS. Performing peptide and/or protein separations using chromatography, electrophoresis or by isolating cellular compartments will help to reduce sample complexity. When designing experiments, it is important to decide what is the smallest subset of proteome that would suit your experiment. For example, to focus on proteins located in the mitochondria, isolate and perform mass spectrometry on the mitochondria only. The mass spectrometer and reverse phase columns have limited loading capacity. By loading the same

protein amount but reducing the range of proteins present in the sample (from >20,000 in a complex whole-cell extract to ~1400 in isolated mitochondria), it is possible to increase the signal for each protein and improve both the proteome coverage and the confidence of peptide identification and quantitation.

USING STABLE ISOTOPES TO ACHIEVE ABSOLUTE QUANTITATION

Stable isotopes can be incorporated into synthetic standards to obtain absolute quantitation. Isotope dilution and related approaches have been used in the small molecule field for decades [20]. A known amount of stable-isotope labeled analog of the compound of interest (internal standard) is spiked into a sample containing the unlabeled compound. The intensity of the unlabeled molecule is compared directly to the intensity of the stable-isotope labeled molecule and the peak ratio calculated. For optimal performance, several concentrations of the internal standard should be measured and a standard curve calculated. Some of the earliest peptide and protein based applications of mass spectrometry for tracking and quantitation exploited enzymatically labeled peptides generated via trypsin ^{18}O-exchange [63], protein quantitation using peptides synthesized using ^{13}C, ^{2}H-labeled amino acids [64] and ^{15}N labeled peptide hormones [65]. Barnidge *et al.* [66] used a deuterium-containing peptide from rhodopsin as an internal peptide standard for determining the absolute amount present in rod outer segments. Taken to its logical extreme, it would be feasible to spike a sample with one or more heavy-isotope labeled synthetic peptide reporters for every protein in the predicted proteome, a strategy dubbed Absolute Quantitation (AQUA)[67]. This methodology can also be exploited to provide absolute quantitation of post-translational modifications.

As an alternative to protein quantitation from a single peptide standard, synthesizing or expressing stable-isotope labeled proteins can generate several peptide standards that can be used even in fractionated samples. In Protein Standard Absolute Quantification, PSAQ, stable-isotope labeled proteins are synthesized *in vitro* and purified to homogeneity before adding to the proteomic sample [68, 69]. Mann *et al.* [70] performed "Absolute SILAC" with internal protein standards using recombinant proteins purified from stable-isotope labeled *E. coli*. Additionally, a single synthesized concatemer protein comprised of peptides from 20 proteins of interest (QconCAT) has been generated to quantify a mixture of proteins [71-74]. These isotope dilution strategies are reviewed in [75].

Taken together, these studies show that the absolute quantitation of peptides and proteins using mass spectrometry is feasible. However, the sequence and identity of the peptide/protein of interest must be known so that the internal standard peptide/protein can be synthesized or isolated. Working sample complexity is limited by practical considerations including the labor expense of generating 100's to 1000's of individual stable-isotope labeled peptides and/or proteins.

HARNESSING THE INFORMATION OBTAINED FROM STABLE-ISOTOPE LABELING

For all methodologies except isobaric methods, the MS1 spectra will contain peptide pairs consisting of an unlabeled and a labeled peptide, representing the peptides that can be quantified. Optimally, the mass spectrometer would recognize these pairs and preferentially select the "light" monoisotopic ion for fragmentation, thereby avoiding background and/or contaminating ions and offsetting the added complexity in the sample. This is particularly important for the analysis of complex mass-tagged samples where the number of peptide pairs far exceeds the number of possible fragmentation scans. In principle, the existing user-defined, data-dependent scanning software provided on commercial mass spectrometers can be adapted to direct the mass spectrometer to flag ions that are separated by a pre-defined mass (mass tag) and subject only these to fragmentation. For example, such a setting is called "mass tag" in Xcalibur software for Orbitraps and FT-ICR mass spectrometers (Thermo Finnigan). However, as of the writing of this review, "mass tag" remains to be fully implemented.

In addition to quantitation, stable-isotope labeling has been used to distinguish contaminants from *bona fide* interactors in immunopurifications (I-DIRT) [76]. Tackett *et al.* grew yeast cells containing an affinity-tagged protein in light SILAC media and control yeast cells in heavy media. After mixing the samples and isolating the affinity tagged protein complex, specific protein interactions were identified by mass spectrometry as a single unlabeled peptide (light), but background contaminant proteins present in both the control (heavy) and affinity-tag protein expressing cells (light) were identified as peptide pairs. Another clever use of stable-isotope quantitation is to examine dynamic protein-protein complexes and protein-DNA complexes [49, 77] by combining affinity purification approaches with stable-isotope tagging. Quantification of component stoichiometry of multiprotein complexes has been performed using a peptide-concatenated standard (PCS) strategy [78]. In this strategy, tryptic peptides suitable for quantification are selected from each component of the multiprotein complex and concatenated into a single synthetic protein, resulting in equimolar amounts of each "heavy" reference peptide. Other uses for stable-isotope labeling include measuring the rate of protein turnover [79] and identifying phosphorylation sites [49].

USE OF STABLE ISOTOPES TO OBTAIN FASTER AND MORE ACCURATE PROTEIN IDENTIFICATION

A complementary advantage of stable-isotope labeling is that when both heavy and light forms are subjected to fragmentation, mass shifts are observed in the MS/MS spectra that facilitate deconvolution and peptide sequence analysis. For the simplest case, where only the carboxyl terminus is labeled as in SILAC using lysine and arginine amino acids or ^{18}O labeling, comparing the two fragmentation patterns or selecting both forms to fragment together flags ions that derive from the carboxyl terminus (y-type ions), as those displaying characteristic mass shifts (e.g. 4 Da) (Fig. **4**). Accordingly, comparison of spectra of labeled and unlabeled peptide fragments allows for assignment of peaks as shifting

or non-shifting, permitting assignment of peaks to one ion series or the other and facilitating *de novo* peptide sequence analysis [80-84]. Peak assignment for validation of peptide identifications obtained by database search has been automated in the Validator software suite [85], which recognizes isotopic peptide pairs from searched MS data and compares their identifications and fragmentation patterns. Because database search algorithms do not utilize the embedded information from comparison of labeled and unlabeled peptides, Validator software provides a direct and independent means to validate peptide identifications from database search algorithms.

CONCLUSIONS

Stable isotopes have become a versatile and useful tool in quantitative mass spectrometry. This review has described chemical, enzymatic and metabolic stable-isotope labeling techniques while highlighting the advantages and disadvantages of each method. A wide variety of sample types can be labeled and analyzed including individual proteins and complexes, biofluids, organelles, bacteria, yeast, mammalian cells and tissues. Absolute quantitation is straightforward for a single protein or a protein complex, but remains cost- and/or labor-prohibitive for complex samples. Instead, a subproteome or a complex cell extract are better suited to relative quantitation where one or more samples are compared to a control sample and fold-change is calculated. In addition to quantitation, stable-isotope labeling can be used to identify components and measure the stoichiometry of protein-protein and protein-DNA complexes, to identify posttranslational modifications and background contamination and to aid in peptide identification and validation.

Modern mass spectrometers are capable of remarkable sensitivity, resolution, reproducibility and speed, so that isotopic experiments simple enough to be amenable to manual analysis can achieve precise quantitation of sub-femtomolar samples. However, many challenges remain that affect the quality of results for more interesting experiments on complex samples, offering pitfalls for experienced and naive users alike. Sadly, no isotopic method is proof to the wide range of artifacts that arise due to biological variation, human error, primitive design and implementation of instrumentation control and poorly executed data analysis software. Confounding the situation, proteomics experiments provide spurious answers side-by-side with highly reliable results, often with no clear distinction among them.

Nonetheless, some common principles apply that will enhance the quality of every experiment. A critical component to stable-isotope labeling is achieving the most complete and consistent labeling feasible as this greatly simplifies downstream data analysis. Decreasing sample complexity to improve peptide statistics for each protein allows high confidence in identification and ready discovery of quantitation artifacts. Although software has come a long way in the last decade, manual validation to the level of visual inspection of mass spectrometry spectra remains a critical step. In summary, stable-isotope labeling for protein quantitation by mass spectrometry remains an emerging technology. Like many other proteomic methods, isotopic labeling is a powerful technique but care must be taken to use appropriate controls, including biological and/or technical replicates, to identify potential problems with labeling, sample handling and/or data analysis.

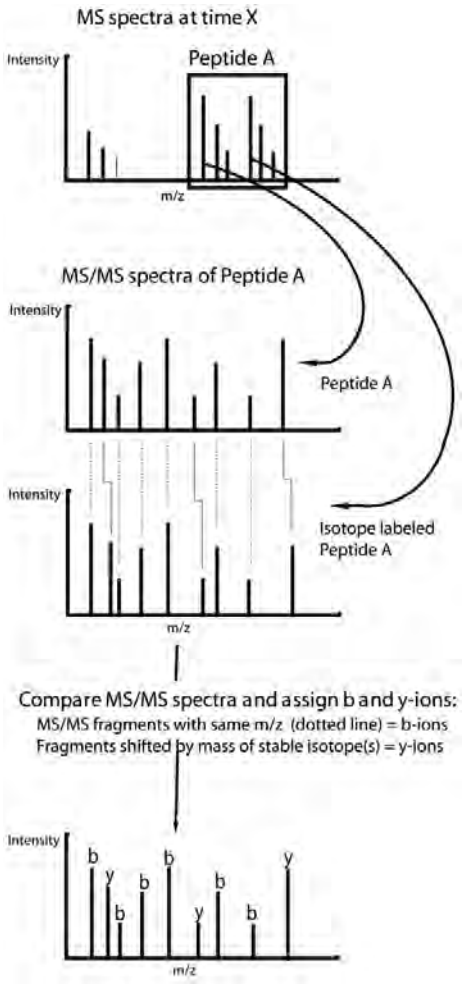

Fig. (4). Identification of b- and y-ions by comparing MS/MS spectra of unlabeled and stable isotope labeled peptides. The top panel shows the MS spectra of peptide A. The middle panels show the MS/MS spectra of unlabeled peptide A and C-terminal stable isotope labeled peptide A. Comparing the fragmentation pattern of the two spectra reveal non-shifting ions (b-ions) and ions that shift by the mass of the stable isotopes (y-ions) and the bottom panel shows the identified b- and y-ions.

ACKNOWLEDGEMENTS

The authors would like to thank Alex Schilling and Juliesta Sylvester for critical reading of the manuscript. K.K. was supported by a BCRP Multidisciplinary Postdoctoral Award W81XWH and S.J.K. was funded by NIH R01 GM60443.

ABBREVIATIONS

AQUA = Absolute QUAntitation peptide strategy

CPAS = Computational Proteomics Analysis System

ESI = ElectroSpray Ionization

ICAT = Isotope-Coded Affinity Tag

ITRAQ = Isobaric Tags for Relative and Absolute Quantitation

MALDI = Matrix-Assisted Laser Desorption Ionization

PSAQ = Protein Standard Absolute Quantification

QconCAT = Q peptide CONCATamers

SILAC = Stable-isotope Labeling with Amino acids in Cell culture

TPP = Trans Proteomic Pipeline

REFERENCES

[1] Aebersold, R. and Mann, M. Mass spectrometry-based proteomics. *Nature*, **2003**, *422* (6928), 198-207.

[2] Fischer, B.; Grossmann, J.; Roth, V.; Gruissem, W.; Baginsky, S. and Buhmann, J.M. Semi-supervised LC/MS alignment for differential proteomics. *Bioinformatics*, **2006**, *22*(14), e132-e140.

[3] Tang, H.; Arnold, R.J.; Alves, P.; Xun, Z.; Clemmer, D.E.; Novotny, M.V.; Reilly, J.P. and Radivojac, P. A computational approach toward label-free protein quantification using predicted peptide detectability. *Bioinformatics*, **2006**, *22*(14), e481-e488.

[4] Ono, M.; Shitashige, M.; Honda, K.; Isobe, T.; Kuwabara, H.; Matsuzuki, H.; Hirohashi, S. and Yamada, T. Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. *Mol. Cell Proteomics*, **2006**, *5*(7), 1338-1347.

[5] Old, W.M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K.G.; Mendoza, A.; Sevinsky, J.R.; Resing, K.A. and Ahn, N.G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell Proteomics*, **2005**, *4*(10), 1487-1502.

[6] Andreev, V.P.; Li, L.; Cao, L.; Gu, Y.; Rejtar, T.; Wu, S.L. and Karger, B.L. A new algorithm using cross-assignment for label-free quantitation with LC-LTQ-FT MS. *J. Proteome Res.*, **2007**, *6*(6), 2186-2194.

[7] Paoletti, A.C.; Parmely, T.J.; Tomomori-Sato, C.; Sato, S.; Zhu, D.; Conaway, R.C.; Conaway, J.W.; Florens, L. and Washburn, M.P. Quantitative proteomic analysis of distinct mammalian Mediator complexes using normalized spectral abundance factors. *Proc. Natl. Acad. Sci. USA*, **2006**, *103*(50), 18928-18933.

[8] Listgarten, J. and Emili, A. Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. *Mol. Cell Proteomics*, **2005**, *4*(4), 419-434.

[9] Wiener, M.C.; Sachs, J.R.; Deyanova, E.G. and Yates, N.A. Differential mass spectrometry: a label-free LC-MS method for finding significant differences in complex peptide and protein mixtures. *Anal. Chem.*, **2004**, *76*(20), 6085-6096.

[10] Wang, G.; Wu, W.W.; Zeng, W.; Chou, C.L. and Shen, R.F. Label-free protein quantification using LC-coupled ion trap or FT mass spectrometry: Reproducibility, linearity, and application with complex proteomes. *J. Proteome Res.*, **2006**, *5*(5), 1214-1223.

[11] Zhang, H.; Yan, W. and Aebersold, R. Chemical probes and tandem mass spectrometry: a strategy for the quantitative analysis of proteomes and subproteomes. *Curr. Opin. Chem. Biol.*, **2004**, *8*(1), 66-75.

[12] Ye, X.; Luke, B.; Andresson, T. and Blonder, J. ^{18}O stable isotope labeling in MS-based proteomics. *Brief Funct. Genomic Proteomic*, **2009**, *8*(2), 136-144.

[13] Ong, S.E. and Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.*, **2005**, *1*(5), 252-262.

[14] Zubarev, R. and Mann, M. On the proper use of mass accuracy in proteomics. *Mol. Cell Proteomics*, **2007**, *6*(3), 377-381.

[15] Liu, T.; Belov, M.; Jaitly, N.; Qian, W. and Smith, R. Accurate mass measurements in proteomics. *Chem. Rev.*, **2007**, *107*(8), 3621-3653.

[16] Mann, M. and Kelleher, N.L. Precision proteomics: the case for high resolution and high mass accuracy. *Proc. Natl. Acad. Sci. USA*, **2008**, *105*(47), 18132-18138.

[17] Yergey, J. A general approach to calculating isotopic distributions for mass spectrometry. *Int. J. Mass Spectr. Ion Phys.*, **1983**, *52*(2-3), 337-349.

[18] Fernandez-de-Cossio, J.; Gonzalez, L.J.; Satomi, Y.; Betancourt, L.; Ramos, Y.; Huerta, V.; Amaro, A.; Besada, V.; Padron, G.; Minamino, N. and Takao, T. Isotopica: a tool for the calculation and viewing of complex isotopic envelopes. *Nucleic Acids Res.*, **2004**, *32*(Web Server issue), W674-W678.

[19] Sykes, M.T. and Williamson, J.R. Envelope: interactive software for modeling and fitting complex isotope distributions. *BMC Bioinformatics*, **2008**, *9*, 446.

[20] Baillie, T.A. The use of stable isotopes in pharmacological research. *Pharmacol. Rev.*, **1981**, *33*(2), 81-132.

[21] Gygi, S.P.; Rist, B.; Gerber, S.A.; Turecek, F.; Gelb, M.H. and Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.*, **1999**, *17*(10), 994-999.

[22] Ji, J.; Chakraborty, A.; Geng, M.; Zhang, X.; Amini, A.; Bina, M. and Regnier, F. Strategy for qualitative and quantitative analysis in proteomics based on signature peptides. *J. Chromatogr. B. Biomed. Sci. Appl.*, **2000**, *745*(1), 197-210.

[23] Yao, X.; Freas, A.; Ramirez, J.; Demirev, P.A. and Fenselau, C. Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus. *Anal. Chem.*, **2001**, *73*(13), 2836-2842.

[24] Oda, Y.; Huang, K.; Cross, F.R.; Cowburn, D. and Chait, B.T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*(12), 6591-6596.

[25] Wu, C.C.; MacCoss, M.J.; Howell, K.E.; Matthews, D.E. and Yates, J.R.3rd. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal. Chem.*, **2004**, *76*(17), 4951-4959.

[26] Ong, S.E.; Kratchmarova, I. and Mann, M. Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC). *J. Proteome Res.*, **2003**, *2*(2), 173-181.

[27] Ong, S.E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D.B.; Steen, H.; Pandey, A. and Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell Proteomics*, **2002**, *1*(5), 376-386.

[28] de Godoy, L.M.; Olsen, J.V.; de Souza, G.A.; Li, G.; Mortensen, P. and Mann, M. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.*, **2006**, *7*(6), R50.

[29] Amanchy, R.; Kalume, D.E.; Iwahori, A.; Zhong, J. and Pandey, A. Phosphoproteome analysis of HeLa cells using stable isotope labeling with amino acids in cell culture (SILAC). *J. Proteome Res.*, **2005**, *4*(5), 1661-1671.

[30] Beynon, R.J. and Pratt, J.M. Metabolic labeling of proteins for proteomics. *Mol. Cell Proteomics*, **2005**, *4*(7), 857-872.

[31] Zhang, R. and Regnier, F.E. Minimizing resolution of isotopically coded peptides in comparative proteomics. *J. Proteome Res.*, **2002**, *1*(2), 139-147.

[32] Zhang, R.; Sioma, C.S.; Wang, S. and Regnier, F.E. Fractionation of isotopically labeled peptides in quantitative proteomics. *Anal. Chem.*, **2001**, *73*(21), 5142-5149.

[33] Zhang, R.; Sioma, C.S.; Thompson, R.A.; Xiong, L. and Regnier, F.E. Controlling deuterium isotope effects in comparative proteomics. *Anal. Chem.*, **2002**, *74*(15), 3662-3669.

[34] Blagoev, B.; Ong, S.E.; Kratchmarova, I. and Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.*, **2004**, *22*(9), 1139-1145.

[35] Molina, H.; Yang, Y.; Ruch, T.; Kim, J.W.; Mortensen, P.; Otto, T.; Nalli, A.; Tang, Q. Q.; Lane, M.D.; Chaerkady, R. and Pandey, A. Temporal profiling of the adipocyte proteome during differentiation using a five-plex SILAC based strategy. *J. Proteome Res.*, **2009**, *8*(1), 48-58.

[36] Griffin, T.J.; Gygi, S.P.; Ideker, T.; Rist, B.; Eng, J.K.; Hood, L. and Aebersold, R. Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae. *Mol. Cell Proteomics*, **2002**, *1*(4), 323-333.

[37] Zhou, H.; Ranish, J.A.; Watts, J.D. and Aebersold, R. Quantitative proteome analysis by solid-phase isotope tagging and mass spectrometry. *Nat. Biotechnol.,* **2002**, *20*(5), 512-515.

[38] Shiio, Y.; Donohoe, S.; Yi, E.C.; Goodlett, D.R.; Aebersold, R. and Eisenman, R.N. Quantitative proteomic analysis of Myc oncoprotein function. *EMBO J.*, **2002**, *21*(19), 5088-5096.

[39] Yan, W.; Lee, H.; Yi, E.C.; Reiss, D.; Shannon, P.; Kwieciszewski, B.K.; Coito, C.; Li, X.J.; Keller, A.; Eng, J.K.; Galitski, T.; Goodlett, D.R.; Aebersold, R. and Katze, M.G. System-based proteomic analysis of the interferon response in human liver cells. *Genome Biol.,* **2004**, *5* (8), R54.

[40] Olsen, J.V.; Andersen, J.S.; Nielsen, P.A.; Nielsen, M.L.; Figeys, D.; Mann, M. and Wisniewski, J.R. HysTag--a novel proteomic quantification tool applied to differential display analysis of membrane proteins from distinct areas of mouse brain. *Mol. Cell Proteomics,* **2004**, *3* (1), 82-92.

[41] Faca, V.; Coram, M.; Phanstiel, D.; Glukhova, V.; Zhang, Q.; Fitzgibbon, M.; McIntosh, M. and Hanash, S. Quantitative analysis of acrylamide labeled serum proteins by LC-MS/MS. *J. Proteome Res*., **2006**, *5*(8), 2009-2018.

[42] Patterson, S.D. From electrophoretically separated protein to identification: strategies for sequence and mass analysis. *Anal. Biochem.,* **1994**, *221*(1), 1-15.

[43] Ross, P.L.; Huang, Y.N.; Marchese, J.N.; Williamson, B.; Parker, K.C.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A. and Pappin, D.J. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell Proteomics,* **2004**, *3*(12), 1154-1169.

[44] Meany, D.L.; Xie, H.; Thompson, L.V.; Arriaga, E.A. and Griffin, T.J. Identification of carbonylated proteins from enriched rat skeletal muscle mitochondria using affinity chromatography-stable isotope labeling and tandem mass spectrometry. *Proteomics*, **2007**, *7*(7), 1150-1163.

[45] Armenta, J.M.; Hoeschele, I. and Lazar, I.M. Differential protein expression analysis using stable isotope labeling and PQD linear ion trap MS technology. *J. Am. Soc. Mass Spectrom.,* **2009**, *20*(7), 1287-1302.

[46] Zhang, Y.; Wolf-Yadlin, A.; Ross, P.L.; Pappin, D.J.; Rush, J.; Lauffenburger, D.A. and White, F. Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules. *Mol. Cell Proteomics,* **2005**, *4*(9), 1240-1250.

[47] Yan, W.; Hwang, D. and Aebersold, R. Quantitative proteomic analysis to profile dynamic changes in the spatial distribution of cellular proteins. *Methods Mol. Biol.,* **2008**, *432*, 389-401.

[48] Ho, J.; Kong, J.W.; Choong, L.Y.; Loh, M.C.; Toy, W.; Chong, P.K.; Wong, C.H.; Wong, C.Y.; Shah, N. and Lim, Y.P. Novel Breast Cancer Metastasis-Associated Proteins. *J. Proteome Res.,* **2009**, *8*(2), 583-594.

[49] Pflieger, D.; Junger, M.A.; Muller, M.; Rinner, O.; Lee, H.; Gehrig, P.M.; Gstaiger, M. and Aebersold, R. Quantitative proteomic analysis of protein complexes: concurrent identification of interactors and their state of phosphorylation. *Mol. Cell Proteomics,* **2008**, *7*(2), 326-346.

[50] Reynolds, K.J.; Yao, X. and Fenselau, C. Proteolytic ^{18}O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent. *J. Proteome Res.,* **2002**, *1*(1), 27-33.

[51] Stewart, I.; Thomson, T. and Figeys, D. ^{18}O labeling: a tool for proteomics. *Rapid Commun. Mass Spectrom,,* **2001**, *15*(24), 2456-2465.

[52] Heller, M.; Mattou, H.; Menzel, C. and Yao, X. Trypsin catalyzed ^{16}O-to-^{18}O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers. *J. Am. Soc. Mass Spectrom.,* **2003**, *14*(7), 704-718.

[53] Mirgorodskaya, O.A.; Kozmin, Y.P.; Titov, M.I.; Korner, R.; Sonksen, C.P. and Roepstorff, P. Quantitation of peptides and proteins by matrix-assisted laser desorption/ionization mass spectrometry using (18)O-labeled internal standards. *Rapid Commun. Mass Spectrom,* **2000**, *14*(14), 1226-1232.

[54] Mueller, L.N.; Brusniak, M.Y.; Mani, D.R. and Aebersold, R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J. Proteome Res.,* **2008**, *7*(1), 51-61.

[55] Lau, K.W.; Jones, A.R.; Swainston, N.; Siepen, J.A. and Hubbard, S.J. Capture and analysis of quantitative proteomic data. *Proteomics*, **2007**, *7*(16), 2787-2799.

[56] Saito, A.; Nagasaki, M.; Oyama, M.; Kozuka-Hata, H.; Semba, K.; Sugano, S.; Yamamoto, T. and Miyano, S. AYUMS: an algorithm for completely automatic quantitation based on LC-MS/MS proteome data and its application to the analysis of signal transduction. *BMC Bioinformatics*, **2007**, *8*, 15.

[57] Pan, C.; Kora, G.; McDonald, W.H.; Tabb, D.; VerBerkmoes, N.C.; Hurst, G.B.; Pelletier, D.A.; Samatova, N.F. and Hettich, R.L. ProRata: A quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation. *Anal. Chem.,* **2006**, *78*(20), 7121-7131.

[58] Bouyssie, D.; Gonzalez de Peredo, A.; Mouton, E.; Albigot, R.; Roussel, L.; Ortega, N.; Cayrol, C.; Burlet-Schiltz, O.; Girard, J.P. and Monsarrat, B. Mascot file parsing and quantification (MFPaQ), a new software to parse, validate, and quantify proteomics data generated by ICAT and SILAC mass spectrometric analyses: application to the proteomics study of membrane proteins from primary human endothelial cells. *Mol. Cell Proteomics,* **2007**, *6*(9), 1621-1637.

[59] Han, D.K.; Eng, J.K.; Zhou, H. and Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.,* **2001**, *19*(10), 946-951.

[60] Li, X.J.; Zhang, H.; Ranish, J.A. and Aebersold, R. Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry. *Anal. Chem.,* **2003**, *75*(23), 6648-6657.

[61] Rauch, A.; Bellew, M.; Eng, J.; Fitzgibbon, M.; Holzman, T.; Hussey, P.; Igra, M.; Maclean, B.; Lin, C.W.; Detter, A.; Fang, R.; Faca, V.; Gafken, P.; Zhang, H.; Whiteaker, J.; Whitaker, J.; States, D.; Hanash, S.; Paulovich, A. and McIntosh, M.W. Computational Proteomics Analysis System (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. J. Proteome Res., **2006**, *5*(1), 112-121.

[62] Cox, J. and Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.,* **2008**, *26*(12), 1367-1372.

[63] Desiderio, D.M. and Kai, M. Preparation of stable isotope-incorporated peptide internal standards for field desorption mass spectrometry quantification of peptides in biologic tissue. *Biomed. Mass Spectrom,,* **1983**, *10*(8), 471-479.

[64] Barr, J.R.; Maggio, V.L.; Patterson, D.G.; Cooper, G.R.; Henderson, L.O.; Turner, W.E.; Smith, S.J.; Hannon, W.H.; Needham, L.L. and Sampson, E.J. Isotope dilution--mass spectrometric quantification of specific proteins: model application with apolipoprotein A-I. *Clin. Chem.,* **1996**, *42*(10), 1676-1682.

[65] Kippen, A.D.; Cerini, F.; Vadas, L.; Stocklin, R.; Vu, L.; Offord, R.E. and Rose, K. Development of an isotope dilution assay for precise determination of insulin, C-peptide, and proinsulin levels in non-diabetic and type II diabetic individuals with comparison to immunoassay. *J. Biol. Chem.,* **1997**, *272*(19), 12513-12522.

[66] Barnidge, D.R.; Dratz, E.A.; Martin, T.; Bonilla, L.E.; Moran, L.B. and Lindall, A. Absolute quantification of the G protein-coupled receptor rhodopsin by LC/MS/MS using proteolysis product peptides and synthetic peptide standards. *Anal. Chem.,* **2003**, *75*(3), 445-451.

[67] Gerber, S.A.; Rush, J.; Stemman, O.; Kirschner, M.W. and Gygi, S.P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. USA,* **2003**, *100*(12), 6940-6945.

[68] Brun, V.; Dupuis, A.; Adrait, A.; Marcellin, M.; Thomas, D.; Court, M.; Vandenesch, F. and Garin, J. Isotope-labeled protein standards: toward absolute quantitative proteomics. *Mol. Cell Proteomics,* **2007**, *6*(12), 2139-2149.

[69] Dupuis, A.; Hennekinne, J.A.; Garin, J. and Brun, V. Protein Standard Absolute Quantification (PSAQ) for improved investigation of staphylococcal food poisoning outbreaks. *Proteomics*, **2008**, *8*(22), 4633-4636.

[70] Hanke, S.; Besir, H.; Oesterhelt, D. and Mann, M. Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *J. Proteome Res.,* **2008**, *7*(3), 1118-1130.

[71] Pratt, J.M.; Simpson, D.M.; Doherty, M.K.; Rivers, J.; Gaskell, S.J. and Beynon, R.J. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.,* **2006**, *1*(2), 1029-1043.

[72] Johnson, H.; Wong, S.C.; Simpson, D.M.; Beynon, R.J. and Gaskell, S.J. Protein quantification by selective isolation and fragmentation of isotopic pairs using FT-ICR MS. *J. Am. Soc. Mass Spectrometry,* **2008**, *19*(7), 973-977.

[73] Rivers, J.; Simpson, D.M.; Robertson, D.H.; Gaskell, S.J. and Beynon, R.J. Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Mol. Cell Proteomics,* **2007**, *6*(8), 1416-1427.

[74] Beynon, R.J.; Doherty, M.K.; Pratt, J.M. and G askell, S.J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods,* **2005**, *2*(8), 587-589.

[75] Brun, V.; Masselon, C.; Garin, J. and Dupuis, A. Isotope dilution strategies for absolute quantitative proteomics. *J. Proteomics,* **2009**, *72*(5), 740-749.

[76] Tackett, A.J.; DeGrasse, J.A.; Sekedat, M.D.; Oeffinger, M.; Rout, M.P. and Chait, B.T. I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. *J. Proteome Res.,* **2005**, *4*(5), 1752-1756.

[77] Mittler, G.; Butter, F. and Mann, M. A SILAC-based DNA protein interaction screen that identifies candidate binding proteins to functional DNA elements. *Genome Res.,* **2009**, *19*(2), 284-293.

[78] Kito, K.; Ota, K.; Fujita, T. and Ito, T. A synthetic protein approach toward accurate mass spectrometric quantification of component stoichiometry of multiprotein complexes. *J. Proteome Res.,* **2007**, *6*(2), 792-800.

[79] Pratt, J.M.; Petty, J.; Riba-Garcia, I.; Robertson, D.H.; Gaskell, S.J.; Oliver, S.G. and Beynon, R.J. Dynamics of protein turnover, a missing dimension in proteomics. *Mol. Cell Proteomics,* **2002**, *1*(8), 579-591.

[80] Hunt, D.F.; Yates, J.R.3rd; Shabanowitz, J.; Winston, S. and Hauer, C.R. Protein sequencing by tandem mass spectrometry. *Proc. Natl. Acad. Sci. USA,* **1986**, *83*(17), 6233-6237.

[81] Schnolzer, M.; Jedrzejewski, P. and L ehmann, W.D. Protease-catalyzed incorporation of 18O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis*, **1996**, *17*(5), 945-953.

[82] Takao, T.; Hori, H.; Okamoto, K.; Harada, A.; Kamachi, M. and Shimonishi, Y. Facile assignment of sequence ions of a peptide labelled with 18O at the carboxyl terminus. *Rapid Commun. Mass Spectrom,* **1991**, *5*(7), 312-315.

[83] Gray, W.R.; Wojcik, L.H. and Futrell, J.H. Application of mass spectrometry to protein chemistry. II. Chemical ionization studies on acetylated permethylated peptides. *Biochem. Biophys. Res. Commun.,* **1970**, *41*(5), 1111-1119.

[84] Rose, K.; Simona, M.G.; Offord, R.E.; Prior, C.P.; Otto, B. and Thatcher, D.R. A new mass-spectrometric C-terminal sequencing technique finds a similarity between gamma-interferon and alpha 2-interferon and identifies a proteolytically clipped gamma-interferon that retains full antiviral activity. *Biochem. J.,* **1983**, *215*(2), 273-277.

[85] Volchenboum, S.L.; Kristjansdottir, K.; Wolfgeher, D. and Kron, S.J. Rapid validation of Mascot search results *via* stable isotope labeling, pair picking, and deconvolution of fragmentation patterns. *Mol. Cell Proteomics,* **2009**, *8*(8), 2011-2022.

Received: October 23, 2009 Revised: February 26, 2010 Accepted: March 04, 2010